

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 29, 2003, 13:42:08 ; Search time 99 seconds
(without alignments)
1099.982 Million cell updates/sec

Title: US-10-005-549-2
Perfect score: 2229
Sequence: 1 MQQPQGGQQGPGGQLGG.....KSLVQIHEKNGWYTPPKEDG 422

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:**
1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mhc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp_vertebrate:**
14: sp_unclassified:**
15: sp_rviro:**
16: sp_bacteriap:**
17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1448.5	65.0	378	11 Q8K228	Q8k228 mus musculus
2	1445.5	64.8	378	11 Q8BUN2	Q8bun2 mus musculus
3	1435	64.4	375	4 Q8WVN8	Q8wvn8 homo sapien
4	1308	58.7	253	11 Q9D7E1	Q9d7e1 mus musculus
5	1191	53.4	230	4 Q9BVX5	Q9bvxs5 homo sapien
6	1149.5	51.6	371	11 Q8BW45	Q8bm45 mus musculus
7	1117	50.1	217	4 Q96WV4	Q96wm4 homo sapien
8	1117	50.1	217	11 Q8K2T0	Q8k2t0 mus musculus
9	1107	49.7	255	11 Q8BVX5	Q8bvxs5 mus musculus
10	1020.5	45.8	397	5 Q9W4Z7	Q9w4z7 drosophila
11	965.5	43.3	394	5 Q46068	Q46068 drosophila
12	841	37.7	471	5 Q93571	Q93571 caenorhabdi
13	808	36.2	217	5 Q8T304	Q8t3u4 drosophila
14	728.5	32.7	199	4 Q96J08	Q96j08 homo sapien
15	440.5	19.8	306	5 Q9VM35	Q9vm35 drosophila
16	364	16.3	71	4 Q9UGL6	Q9ugl6 homo sapien

17	343	15.4	71	4 Q8N4G6	Q8n4g6 homo sapien
18	140	6.3	425	5 Q95039	Q95039 paramecium
19	138	6.2	829	4 Q8N8K6	Q8n8k6 homo sapien
20	136.5	6.1	199	5 Q8T0Z6	Q8t0z6 bombyx mori
21	135	6.1	199	5 P91633	P91633 drosophila
22	133	6.0	1300	12 Q36421	Q36421 alcelaphine
23	131.5	5.9	426	4 Q12937	Q12937 homo sapien
24	131.5	5.9	501	4 Q8WNB4	Q8wnb4 homo sapien
25	131.5	5.9	523	4 Q99932	Q99932 homo sapien
26	128.5	5.8	154	5 Q8T972	Q8t972 drosophila
27	128	5.7	658	11 Q9JL61	Q9jl61 mus musculus
28	128	5.7	711	4 Q00301	Q00301 homo sapien
29	127.5	5.7	207	4 Q969F4	Q969f4 homo sapien
30	127.5	5.7	207	11 Q91X63	Q91x63 mus musculus
31	127	5.7	115	4 Q9UQH5	Q9uqh5 mus musculus
32	127	5.7	342	5 Q61580	Q61580 homo sapien
33	127	5.7	525	11 Q8BIY1	Q8biy1 mus musculus
34	126	5.7	209	13 Q8AW04	Q8aw04 brachydanio
35	126	5.7	430	4 Q9H6U9	Q9h6u9 homo sapien
36	125.5	5.6	1256	11 Q99M76	Q99m76 rattus norv
37	125	5.6	1152	4 Q92603	Q92603 homo sapien
38	125	5.6	1204	4 Q00211	Q00211 homo sapien
39	125	5.6	1235	4 Q9H2G2	Q9h2g2 homo sapien
40	123.5	5.5	964	16 Q8FPA7	Q8fpa7 corynebacte
41	122.5	5.5	327	16 Q93IV2	Q93iv2 streptomyce
42	122.5	5.5	1613	11 Q8K1P7	Q8kip7 rattus norv
43	122	5.5	335	5 Q8SZS4	Q8szs4 drosophila
44	122	5.5	376	5 Q9W2X5	Q9w2x5 drosophila
45	121.5	5.5	147	10 Q957H6	Q957h6 guillardia

ALIGNMENTS

RESULT 1

ID	Q8K228	PRELIMINARY;	PRT;	378 AA.
AC	Q8K228			
DT	01-OCT-2002 (Tremblrel. 22, Created)			
DT	01-OCT-2002 (Tremblrel. 22, Last sequence update)			
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE	Similar to NICE-5 protein (EC 6.3.2.19) (Ubiquitin-conjugating enzyme			
DE	-E2) (Ubiquitin-protein ligase) (Ubiquitin carrier protein).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER			
CC	PROTEINS (BY SIMILARITY).			
CC	-!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +			
CC	DIPHOSPHATE + PROTEIN N-UBIQUITYLYLSINE.			
CC	-!- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.			
CC	-!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-			
CC	THIOLESTER FORMATION (BY SIMILARITY).			
CC	-!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.			
DR	EMBL: BC029111; AAH29111.1; -			
DR	InterPro: IPR000608; UBQ_conjugat.			
DR	Fam; PF001179; UQ_con; 1.			
DR	ProDom; PD000461; UBQ_conjugat; 1.			
DR	SMART; SM00212; UBCC; 1.			
DR	PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.			
KW	Ligase; Ubl conjugation pathway.			
SQ	SEQUENCE 378 AA; 42923 MW; D1690A9C4BC6D8DC CRC64;			

Query Match 65.0%; Score 1448.5; DB 11; Length 378;
Best Local Similarity 74.0%; Pred. No. 1.7e-96;
Matches 285; Conservative 33; Mismatches 52; Indels 15; Gaps 4;

QY 41 LRRLKLESIFHRGHERFRIASACLDLSCEFLLAGAGAGAGAPGPHLPGRSV-PG 99

```

Db      6 LKAEKFLASIFDKNHERFRIVSWKLDLHCQFLV-----PPPPPPGSSLSPP 54
QY     100 DPVRIHCNITESPVPPIWHSVESDDPNLAALVRLVDIKGNTLLQLHLKRIISDLCKL 159
Db      55 PPLTLHCNITESPSSPIWFDSDDPNLTSLVERLED-TKNSSSLRQLKWLICDLCL 113
QY     160 YNLPQHPDVMELDQPLPAEQ--CTQEDVSSDEDEEMPEDETDLDHYEMKEEPEAGKKS 217
Db      114 YNLPKHLVDVMDLQPLTGQGTTEVTSEEEEEEAEEDIEDLDHYEMKEEPEAGKKS 173
QY     218 EDDGIGKENLAILEKIKKNQODYLNAGVSGSVQATDRLMKELRDYRSQSFKGNYAVE 277
Db      174 EDEGIEKENLAILEKIRKTQODHLNGAVSGSVQASDRLMKELRDYRSQYKAGIYSVE 233
QY     278 LVNDSLDYDWNVKKLVKDDQSDSALHNDLQILKEKEGADFIILLNFSKDNFPDPFVRVSP 337
Db      234 LTNDSLYDWNVKKLVKDDQSDSALHNDLQILKEKEGADFIILLNFSKDNFPDPFVRVSP 293
QY     338 VLSGGVYLGGAICMELLTKQGWSSAYSIESVIMQISATLVKGVKARVQFGANKSOYSLTR 397
Db      294 VLSGGVYLGGAICMELLTKQGWSSAYSIESVIMQINATLVKGVKARVQFGANKQYNLAR 353
QY     398 AQSYSKSLVQIHEKNGWYTPPKEDG 422
Db      354 AQSYSNSIVQIHEKNGWYTPPKEDG 378

```

RESULT 2

```

Q8WBUN2 PRELIMINARY; .PRT; 378 AA.
AC Q8WBUN2;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Similar to NICE-5 protein homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
DR EMBL; AK083216; BAC38813.1; -.
SQ SEQUENCE 378 AA; 42950 MW; 4A81CA85400A1313 CRC64;

```

```

Query Match 64.8%; Score 1445.5; DB 11; Length 378;
Best Local Similarity 73.8%; Pred. No. 2.8e-96;
Matches 284; Conservative 34; Mismatches 52; Indels 15; Gaps 4;

QY     41 LRRELKLLSFHGRHERFRITASACLDLSCEFLLAGAGAGAGAGAPGPHLPGRGVP 99
Db      6 LKAEKFLASIFDKNHERFRIVSWKLDLHCQFLV-----PPPPPPGSSLSPP 54
QY     100 DPVRIHCNITESPVPPIWHSVESDDPNLAALVRLVDIKGNTLLQLHLKRIISDLCKL 159
Db      55 PPLTLHCNITESPSSPIWFDSDDPNLTSLVERLED-TKNSSSLRQLKWLICDLCL 113
QY     160 YNLPQHPDVMELDQPLPAEQ--CTQEDVSSDEDEEMPEDETDLDHYEMKEEPEAGKKS 217
Db      114 YNLPKHLVDVMDLQPLTGQGTTEVTSEEEEEEAEEDIEDLDHYEMKEEPEAGKKS 173
QY     218 EDDGIGKENLAILEKIKKNQODYLNAGVSGSVQATDRLMKELRDYRSQSFKGNYAVE 277
Db      174 EDEGIEKENLAILEKIRKTQODHLNGAVSGSVQASDRLMKELRDYRSQYKAGIYSVE 233
QY     278 LVNDSLDYDWNVKKLVKDDQSDSALHNDLQILKEKEGADFIILLNFSKDNFPDPFVRVSP 337

```

```

Db      234 LVNDSLDYDWNVKKLVKDDQSDSALHNDLQILKEKEGADFIILLNFSKDNFPDPFVRVSP 293
QY     338 VLSGGVYLGGAICMELLTKQGWSSAYSIESVIMQISATLVKGVKARVQFGANKSOYSLTR 397
Db      294 VLSGGVYLGGAICMELLTKQGWSSAYSIESVIMQINATLVKGVKARVQFGANKQYNLAR 353
QY     398 AQSYSKSLVQIHEKNGWYTPPKEDG 422
Db      354 AQSYSNSIVQIHEKNGWYTPPKEDG 378

RESULT 3
Q8WBUN8 PRELIMINARY; .PRT; 375 AA.
AC Q8WBUN8;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to NICE-5 protein (EC 6.3.2.19) (Ubiquitin-conjugating enzyme
DE E2) (Ubiquitin-protein ligase) (Ubiquitin carrier protein).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +
CC DIPHOSPHATE + PROTEIN N-UBIQUITYLYSINE.
CC -1- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.
CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR EMBL; BC017708; AAH17708.1; -.
DR InterPro; IPR006575; RWD.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UBQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00591; RWD; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
KW Ligase; Ub1 conjugation pathway.
SQ SEQUENCE 375 AA; 42818 MW; 7DE07315E89178A3 CRC64;

```

```

Query Match 64.4%; Score 1435; DB 4; Length 375;
Best Local Similarity 74.1%; Pred. No. 1.6e-95;
Matches 286; Conservative 33; Mismatches 47; Indels 20; Gaps 6;

QY     41 LRRELKLLSFHGRHERFRITASACLDLSCEFLLAGAGAGAGAGAPGPHLPGRGVP 99
Db      6 LKAEKFLASIFDKNHERFRIVSWKLDLHCQFLVPOQG-----SPHSLEPP----- 51
QY     100 DPVRIHCNITESPVPPIWHSVESDDPNLAALVRLVDIKGNTLLQLHLKRIISDLCKL 159
Db      52 -PPLTLHCNITESPSSPIWFDSDDPNLTSLVERLED-TKNNNLRQLKWLICELCSL 109
QY     160 YNLPQHPDVMELDQPLPAEQ-CTQEDVSS--EDEDEEMPEDETDLDHYEMKEEPEAGK 216
Db      110 YNLPKHLVDVMDLQPLTGQGTTEVTSEEEEEEAEEDIEDLDHYEMKEEPEISGKK 169
QY     217 SEDDGIGKENLAILEKIKKNQODYLNAGVSGSVQATDRLMKELRDYRSQSFKGNYAV 276
Db      170 SEDEGIEKENLAILEKIRKTQODHLNGAVSGSVQASDRLMKELRDYRSQYKGTIYS 229
QY     277 LVNDSLDYDWNVKKLVKDDQSDSALHNDLQILKEKEGADFIILLNFSKDNFPDPFVRVSP 336
Db      230 ELINDSLYDWHVKKLVKDDQSDSALHNDLQILKEKEGADFIILLNFSKDNFPDPFVRVWL 289

```

```

OY 337 PVLGGVYLGGAICMELLTQGWSSAYSIESVIMQISATLVKGRVQFGANKSOYSLT 396
DB 290 PVLGGVYLGGAICMELLTQGWSSAYSIESVIMQISATLVKGRVQFGANKSOYSLT 349
OY 397 RAQGSYKSLVQIHKEKNGWYTPPKEDG 422
DB 350 RAQGSYNSIVQIHKEKNGWYTPPKEDG 375

RESULT 4
Q9D7E1 PRELIMINARY; PRT; 253 AA.
AC Q9D7E1
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Adult male tongue cDNA, RIKEN full-length enriched library,
DE clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA Segabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20530913; PubMed=11076861;

```

```

RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Suni M., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
DR EMBL; AK009324; BAB26217.2;
SQ SEQUENCE 253 AA; 28487 MW; 8FB9519FF26585CF CRC64;

Query Match 58.7%; Score 1308; DB 11; Length 253;
Best Local Similarity 99.6%; Pred. No. 1.4e-86;
Matches 252; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 170 MLDQPLPAEQCTQEDYSSDEDEMPEDTDLHYEMKEEPAEGKSDGIGKENLAI 229
DB 1 MLDQPLPAEQCTQEDYSSDEDEMPEDTDLHYEMKEEPAEGKSDGIGKENLAI 60
OY 230 LEKIKKNQODYLNQAVSGSVQATDRLMKELRIYRSQSPKGNAYVELVNDSLYDNVYK 289
DB 61 LEKIKKNQODYLNQAVSGSVQATDRLMKELRIYRSQSPKGNAYVELVNDSLYDNVYK 120
OY 290 LLKVDQDSALHNDLQILKEGADFIILNFSKDNFPDPFPFVVRVSPVLGGYVGGGA 349
DB 121 LLKVDQDSALHNDLQILKEGADFIILNFSKDNFPDPFPFVVRVSPVLGGYVGGGA 180
OY 350 ICMLTLTKGWSAYSIESVIMQISATLVKGRVQFGANKSOYSLTRAQOQSKSLVQIH 409
DB 181 ICMLTLTKGWSAYSIESVIMQISATLVKGRVQFGANKSOYSLTRAQOQSKSLVQIH 240
OY 410 EKNGWYTPPKEDG 422.
DB 241 EKNGWYTPPKEDG 253

RESULT 5
Q9BVX5 PRELIMINARY; PRT; 230 AA.
AC Q9BVX5
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical ubiquitin-protein ligase (EC 6.3.2.19) (Ubiquitin-
DE conjugating enzyme E2) (Ubiquitin carrier protein)
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE -> AMP +
CC DIPHOSPHATE + PROTEIN N-UBIQUITYLYSINE.
CC -1- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.
CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR EMBL; BC000848; AA000848.1;
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UBQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
KW Hypothetical protein; Ligase; Ub1 conjugation pathway.
FT NON_TER 1

```



```

AC O8K2T0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to EG:2588.2 gene product (EC 6.3.2.19) (Ubiquitin-conjugating
DE enzyme E2) (Ubiquitin-protein ligase) (Ubiquitin carrier
DE protein).
GN 2310012M18RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +
CC DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.
CC -!- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.
CC -!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR EMBL; BC030044; AAH30044.1; -.
DR MGD; MGI:1917343; 2310012M18RIK.
DR InterPro; IPR000608; UBO_conjugat.
DR Pfam; PF00179; UQ_con; 1.
DR PRODOM; PD000461; UBO_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.
DR KW Ligase; ubi conjugation pathway.
SQ SEQUENCE 217 AA; 24277 MW; 3C74AF937AE2BCCE CRC64;

Query Match 50.1%; Score 1117; DB 11; Length 217;
Best Local Similarity 100.0%; Pred. No. 6.8e-73;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 MKEEPAGKSEDDGIGKENLAILEIKKKNORODYLVGSGVQATDRLMKELRIYR 265
DB 1 MKEEPAGKSEDDGIGKENLAILEIKKKNORODYLVGSGVQATDRLMKELRIYR 60

QY 266 SQSPKGGNYAVELVNDLSYDNVKKLVQDQSDALHNDLIQLKEGADFIILNFSKDNF 325
DB 61 SQSPKGGNYAVELVNDLSYDNVKKLVQDQSDALHNDLIQLKEGADFIILNFSKDNF 120

QY 326 PFDPFVRVSPVLGGYVGGGAIICMELLTKQWSSAYSIESVIMQISATLVKGRARVQ 385
DB 121 PFDPFVRVSPVLGGYVGGGAIICMELLTKQWSSAYSIESVIMQISATLVKGRARVQ 180

QY 386 FGANKSQSLTRAQGSYKSLVQIHEKNGWYTPPKEDG 422
DB 181 FGANKSQSLTRAQGSYKSLVQIHEKNGWYTPPKEDG 217

RESULT 9
Q8BVX5 PRELIMINARY; PRT; 255 AA.
AC Q8BVX5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Similar to NICE-5 protein homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,

```

```

RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK076148; BAC36218.1; -.
SQ SEQUENCE 255 AA; 28916 MW; 2F5077B3FAC2D6DF CRC64;

Query Match 49.7%; Score 1107; DB 11; Length 255;
Best Local Similarity 83.1%; Pred. No. 4.5e-72;
Matches 212; Conservative 20; Mismatches 21; Indels 2; Gaps 1;

QY 170 MLDQPLPAEQ-CTQEDVSEDEDEEMPEDELDHYEMKEEPAEKGKSEDDGIGKENL 227
DB 1 MLDQPLPTGQGTTEVTSEEEEAEDIEDLDHYEMKEEPIKNGKSEDEGIEKENL 60

QY 228 AILEKIKKNORODYLVGSGVQATDRLMKELRIYRSQSPKGGNYAVELVNDLSYDNV 287
DB 61 AILEKIKKNORODYLVGSGVQATDRLMKELRIYRSQSPKGGNYAVELVNDLSYDNV 120

QY 288 VKLLKVDQSDALHNDLIQLKEGADFIILNFSKDNFDPFVRVSPVLGGYVGG 347
DB 121 VKLLKVDQSDALHNDLIQLKEGADFIILNFSKDNFDPFVRVSPVLGGYVGG 180

QY 348 GAICMELLTKQWSSAYSIESVIMQISATLVKGRARVQFGANKSQSLTRAQGSYKSLVQ 407
DB 181 GAIICMELLTKQWSSAYSIESVIMQINATLVKGRARVQFGANKSQSLTRAQGSYKSLVQ 240

QY 408 IHEKNGWYTPPKEDG 422
DB 241 IHEKNGWYTPPKEDG 255

RESULT 10
Q9W4Z7 PRELIMINARY; PRT; 397 AA.
AC Q9W4Z7; Q9W4Z8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 23, Last annotation update)
DE EG:2588.2 protein (EC 6.3.2.19).
GN EG:2588.2 OR CG2924.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Cherry J.M., Cawley S., Dahlke C., Cadiou E., Center A., Chandra I.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor S.P., Gabor A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hristov N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodish C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

```

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Remington K., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhang M., Zhang G., Zhao Q., Zheng L.,
RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +
CC DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.
CC -1- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
CC EMBL; AE003423; AAF45767.1; -
DR EMBL; AE003423; AAF45768.1; -
DR FlyBase; FBgn0023528; EG:25E8.2.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UBQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.
KW Hypothetical protein: Alternative splicing; Ligase;
KW Ubl conjugation pathway.
FT VARSPLIC 1 180 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 397 AA; 44391 MW; 37081F41680C9D9B CRC64;

Query Match 45.8%; Score 1020.5; DB 5; Length 397;
Best Local Similarity 49.0%; Pred. No. 1.5e-61;
Matches 202; Conservative 72; Mismatches 87; Indels 51; Gaps 4;

QY 41 LRRELKLESTFHRGHERFRIASACDLDELCEFLLAGAGGAGAGAPGPHLPGRGSPGD 100
Db 7 LKQETKLEKIFPKNHERFQILNSVDLLCRFI-----DKNKG 45

QY 101 PVRIHCNITESYPAPPPITWSVESDDPNLAALVRLVDIKKNTLLLOHLKRIISDLCKLY 160
Db 46 RYDIHANITETYPSPVPVFAESETSVTNAVQILSNTRNGRDNHVNQVIGILLRELCLRH 105

QY 161 NLPQHPDVMELDQPLPAEQCTQEDVSEDE-----DEEMPEDELDHYEM 206
Db 106 NVLPPLPDIDNALPLQTPPPSPASPLRCEQRPGGGAGGGGPHGNEETSDQEIPIG 165

QY 207 KEPEPAEGKK-----SEDDGIGKENLAILEKIKKQRODYLANGVSGSVQATD 254
Db 166 ESEQESEDDELPLEMDVRSITSKDDMEVEHLATLEKLRQSQDYDLKGVSGSVQATD 225

QY 255 RLKMLRLDIYSQSKGNYAVELVNDSDYDNVVKLLKYDDQSALHNDLQILKEKGAADF 314
Db 226 RLKMLRLDIYSQSKGNYAVELVNDSDYDNVVKLLKYDDQSALHNDLQILKEKGAADF 285

QY 315 ILLNFSKDNFPDPFPVVRVSVPLSGGYVGGGALCMELTCKQGWSSAYSIESVIMQIS 374
Db 286 ILLNLFKETYPTPEPPFVRVVRVHPIISGGYVIGGALCMELTCKQGWSSAYSIESVIMQIS 345

QY 375 ATLKVGKARVQFGANKS-----QYSLTRAQSKYKSLVQIHEKNGVTPPKEDG 422
Db 346 ATLKVGKARVQFGANKS-----QYSLTRAQSKYKSLVQIHEKNGVTPPKEDG 397

RESULT 11
046068

RESULT 12

ID 046068 PRELIMINARY; PRT; 394 AA.
AC 046068;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE EG:25E8.2 protein (EC 6.3.2.19) (Ubiquitin-conjugating enzyme E2)
DE (Ubiquitin-protein ligase) (Ubiquitin carrier protein).
GN EG:25E8.2 OR C02924.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; *Drosophila*.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy L., Harris D., Barrell B.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +
CC DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.
CC -1- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.
CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR EMBL; AL009196; CAA15712.1; -
DR FlyBase; FBgn0023528; EG:25E8.2.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UBQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.
KW Ligase; Ubl conjugation pathway.
SQ SEQUENCE 394 AA; 44000 MW; 83E2AFB825D6CD76 CRC64;

Query Match 43.3%; Score 965.5; DB 5; Length 394;
Best Local Similarity 48.1%; Pred. No. 1.4e-61;
Matches 194; Conservative 71; Mismatches 87; Indels 51; Gaps 4;

QY 41 LRRELKLESTFHRGHERFRIASACDLDELCEFLLAGAGGAGAGAPGPHLPGRGSPGD 100
Db 7 LKQETKLEKIFPKNHERFQILNSVDLLCRFI-----DKNKG 45

QY 101 PVRIHCNITESYPAPPPITWSVESDDPNLAALVRLVDIKKNTLLLOHLKRIISDLCKLY 160
Db 46 RYDIHANITETYPSPVPVFAESETSVTNAVQILSNTRNGRDNHVNQVIGILLRELCLRH 105

QY 161 NLPQHPDVMELDQPLPAEQCTQEDVSEDE-----DEEMPEDELDHYEM 206
Db 106 NVLPPLPDIDNALPLQTPPPSPASPLRCEQRPGGGAGGGGPHGNEETSDQEIPIG 165

QY 207 KEPEPAEGKK-----SEDDGIGKENLAILEKIKKQRODYLANGVSGSVQATD 254
Db 166 ESEQESEDDELPLEMDVRSITSKDDMEVEHLATLEKLRQSQDYDLKGVSGSVQATD 225

QY 255 RLKMLRLDIYSQSKGNYAVELVNDSDYDNVVKLLKYDDQSALHNDLQILKEKGAADF 314
Db 226 RLKMLRLDIYSQSKGNYAVELVNDSDYDNVVKLLKYDDQSALHNDLQILKEKGAADF 285

QY 315 ILLNFSKDNFPDPFPVVRVSVPLSGGYVGGGALCMELTCKQGWSSAYSIESVIMQIS 374
Db 286 ILLNLFKETYPTPEPPFVRVVRVHPIISGGYVIGGALCMELTCKQGWSSAYSIESVIMQIS 345

QY 375 ATLKVGKARVQFGANKS-----QYSLTRAQSKYKSLVQIHEKNG 413
Db 346 ATLKVGKARVQFGANKS-----QYSLTRAQSKYKSLVQIHEKNG 388

<hr/>					
Q93571	PRELIMINARY;	PRT;	471 AA.		
ID AC	Q93571;				
DT DT	01-FEB-1997 (TREMBlrel. 02, Created)				
DT DT	01-FEB-1997 (TREMBlrel. 02, Last sequence update)				
DT DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)				
DE DE	F53H2.8 protein (EC 6.3.2.19) (Ubiquitin-conjugating enzyme E2)				
DE GN	(Ubiquitin-protein ligase) (Ubiquitin carrier protein), F53H2.8.				
OS OS	Caenorhabditis elegans.				
OC OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;				
OC OC	Rhabditiidae; Pladeriniae; Caenorhabditis.				
OX NX	NCBI_TaxID=6239;				
RP RP	[1]				
RP RP	SEQUENCE FROM N.A.				
RA RA	Wilkinson J.;				
RL RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.				
RN RN	[2]				
RX RX	SEQUENCE FROM N.A.				
RX RX	MEDLINE=99069613; PubMed=9851916;				
RT RT	none;				
RT RT	"Genome sequence of the nematode C.elegans: A platform for				
RT RT	investigating biology.";				
RL RL	Science 282:2012-2018(1998).				
CC CC	-1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER				
CC CC	PROTEINS (BY SIMILARITY).				
CC CC	-1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +				
CC CC	DIPHOSPHATE + PROTEIN N-BIQUITVLLYSINE.				
CC CC	-1- PATHWAY: UBIQUITIN CONJUGATION; SECOND STEP.				
CC CC	-1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-				
CC CC	THIOESTER FORMATION (BY SIMILARITY).				
CC CC	-1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.				
DR DR	EMBL; Z79754; CAB02096.1; --				
DR DR	WormPep; F25H2.8; CE09653.				
DR DR	InterPro; IPR002048; EF-hand.				
DR DR	InterPro; IPR000608; UBQ_conjugat.				
DR DR	Pfam; PF001179; UQ_con; 1.				
DR DR	ProDom; PD000461; UBQ_conjugat; 1.				
DR DR	SMART; SMO0212; UBCc; 1.				
DR DR	PROSITE; PS00018; EF_HAND; 1.				
DR DR	PROSITE; PSS0127; UBIQUITIN_CONJUGAT_2; 1.				
KW KW	Ligase; ubl conjugation pathway.				
SQ SQ	SEQUENCE 471 AA; 54029 MW; 69EEBCAA5B510D78 CRC64;				
<hr/>					
Query Match	37.7%;	Score	841;	DB	5;
Best Local Similarity	44.9%;	Pred. No.	1.7e-52;		
Matches	180;	Conservative	69;	Mismatches	100;
				Indels	52;
				Gaps	10;
<hr/>					
QY QY	40 CLRR--ELKLLSEIFHRHERFRIASCLDELSCFEFLLAGAGGAGAGAAPGLPPRGS	96			
Db Db	: : : : : : : : : : : : :				
	3 CLRKLAKDIOVLKFLPKNNHFOLLSASVDLSMKFTNAENKG-----	46			
<hr/>					
QY QY	97 VPGDPVRHCHNTESYPAVPPIINSVEDD-PNLAAVLERLVDIRKGNWLLLOHLKRISD	155			
Db Db	: : : : : : : : : : : : :				
	47 -----IVVTANIQENYPQPPIPFSESDPDVIGMSQLRTETEE-STNILHQVHRLVSD	100			
<hr/>					
QY QY	156 LCKLYLN-----PQHPDY-----EMLDQPLPAEQCTQEDVSSEDDEEMP	195			
Db Db	: : : : : : : : : : : : :				
	101 LCSFYNLQMPCLPQAPPYRDDIDEGRSDISTTSEPIDDDMAGGEVDDDEEREEDD	160			
<hr/>					
QY QY	196 EDTE-DLDHYEMEKEEPAGEKKSEDDGIGKENLAILEKIKKNRODYINGAVSGVQATD	254			
Db Db	: : : : : : : : : : : : :				
	161 EDADGDIEIVEMAEDPT---SQHDVGVSKEGLDMDKVSKINQQHLDGKVGQSITATD	217			
<hr/>					
QY QY	255 RLMAKELRDIVRSOFSGNGYAVEL-VNDSLYDNWVKLLKYDDDSALHNDLQILKEKGAD	313			
Db Db	: : : : : : : : : : : : :				
	218 RLMAKELRDIVRSOFSGNGYAVEL-VNDSLYDNWVKLLKYDDDSALHNDLQILKEKGAD	313			
<hr/>					
QY QY	314 FILLNFSEFKONFPDPFVRVSVLGGGVVLGGGAICMELLTKQGWSAYSIESTVMQI	373			
Db Db	: : : : : : : : : : : : :				
	278 HLLFSTFNKFCDCPPFVRVAPHINQGFVLGGGAICMELLTKQGWSAYSIESTLQI	337			
<hr/>					
QY QY	374 SATLVKRGARVFQA-NKSOYLSLTRAQQSYKSLVOIHCKNG	413			

Search completed: July 29, 2003, 13:47:10
Job time : 102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 14:10:51 ; Search time 10661 seconds
(without alignments)
8649.310 Million cell updates/sec

Title: US-10-005-549-1

Perfect score: 2254

Sequence: 1 ctoctctctctctacttg.....tttgaaaaaaaaaaaaa 2254

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1790.4	79.4	3230	9	AY112698 Homo sapi
2	1632.4	73.3	1880	6	AX017982 Sequence
3	1652.4	73.3	1880	6	BD137288 Human nuc
4	1548.4	68.7	3094	10	AY112699 Mus muscu
5	1330	59.0	1330	6	AX017833 Sequence
6	1330	59.0	1330	6	BD137243 Human nuc
7	1160.4	51.5	1162	9	BC000848 Homo sapi
8	1115.4	49.5	2523	6	AX171419 Sequence
9	1115.4	49.5	2523	9	AK056388 Homo sapi
10	998	44.3	1173	10	BC030044 Mus muscu
11	969.4	43.0	992	9	BC009286 Homo sapi
12	880.8	39.1	885	9	HS24243656
13	794	35.2	808	17	AF116721 Homo sapi
14	772	34.3	772	9	BC015316 Homo sapi
15	681.8	30.2	853	10	BC051487 Mus muscu
16	567	25.2	184956	2	AC144817 Mus muscu
17	565.6	25.1	2782	10	BC029111 Mus muscu
18	557	24.7	2939	9	BC017708 Homo sapi
19	557	24.7	2969	9	HS24243656
20	488.4	21.7	64759	9	AL592078 Human DNA
21	478.2	21.2	239353	2	AC128996 Rattus no
22	478.2	21.2	249982	2	AC108661 Rattus no
23	461.2	20.5	2988	6	AX405769 Sequence
24	461.2	20.5	2993	9	AK000617 Homo sapi
25	442.8	19.6	62727	10	AL808107 Mouse DNA
26	432	18.2	432	6	AX410956 Sequence
27	422.6	18.7	5281	9	HSU08191 Human R kap
28	407.6	18.1	222697	2	AC094121 Rattus no
29	407.6	18.1	225763	2	AC115241 Rattus no
30	407.6	18.1	236356	2	AC128343 Rattus no
31	406	18.0	190148	2	AC102392 Mus muscu
32	388	17.2	240986	2	AC140674 Mus muscu
33	374.8	16.6	190148	2	AC102392 Mus muscu
34	371.4	16.5	585	9	AF397158 Homo sapi
35	371.4	16.5	870	6	AX384879 Sequence
36	371.4	16.5	2235	9	HUMGEA R03432 Human 18S r
37	371.4	16.5	25606	2	AC010554 Homo sapi
38	371.4	16.5	42999	6	AX664499 Sequence
39	371.4	16.5	42999	9	HSU13369 Human ribos
40	371.4	16.5	43962	2	AC068192 Homo sapi
41	371.4	16.5	61490	2	AC025630 Homo sapi
42	371.4	16.5	85098	2	AL158197 Homo sapi
43	371.4	16.5	112833	2	AC068881 Homo sapi
44	371.4	16.5	136630	2	AC023572 Homo sapi
45	371.4	16.5	140653	2	AL355134 Homo sapi

ALIGNMENTS

RESULT 1
AY112698
LOCUS
DEFINITION Homo sapiens ubiquitin-conjugating enzyme E2Q (UBE2Q) mRNA,
complete cds.
ACCESSION AY112698
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3230)
AUTHORS Altmann,M.E., Schulze,E., Adham,I., Koehler,M. and Engel,W.
TITLE Isolation and characterization of the human UBE2Q gene and its

murine ortholog					
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3230)				
AUTHORS	Altman, M. E.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-May-2002) University Goettingen, Human Genetics, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany				
FEATURES	Location/Qualifiers				
	1..3230				
source	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
gene	/chromosome="1"				
	/map="1q21-q23"				
CDS	1..3230				
	/gene="UBE2Q"				
	67..1335				
	/gene="UBE2Q"				
	/codon_start=1				
	/product="ubiquitin-conjugating enzyme E2Q"				
	/protein_id="AA060814.1"				
	/db_xref="GI:31321953"				
	/translation="MQQPQPGQQPQPGQOLGQGAAPGAGGGPGGPGPGLRRE LKLSEIFRHEFRASACDELSEFLLAGAGAGAGAAPGPHPPRGSPGDPV RIHNTESYPAPVPIWVSDESDPNLAVERLYDIRKNTLLLOHLKRIISDLCKLY NLQPHDYEMLDQPLPAEQCTQEDVSEDEDEPTEIDLDHYEMKEEPFAEGKKE DDGIGKENLAILEIKKNORODYLVSGVQATRLMKELRDIYRSQSPKGNVAV ELVNDLSLDVNVKLLKYDODSALHNDLQILKEKRGADPILLNFSEKDNFPDPFVAV VSPVLGGVVLGGGAICMELLTKGWSNYSIESVIMQISNLYVKRQVFGANKSQ YSLTRAQOYSKLSVQIHEKNWYTPPKEDG"				
	BASE COUNT 826 a 770 c 833 g 801 t				
ORIGIN					
Query Match		79.4%; Score 1790.4; DB 9; Length 3230;			
Best Local Similarity		99.9%; Pred. No. 4.4e-212;			
Matches 1791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	451	GGCGAGCCCGGAGCGAGCGGAGGCTCCGGCCCGCCCGCGGAGCGGAGCGGAG 510			
DB	1	GGCGAGCCCGGAGCGAGCGGAGGCTCCGGCCCGCCCGCGGAGCGGAGCGGAG 60			
QY	511	CGGAGGATGACAGCCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 570			
DB	61	CGGAGGATGACAGCCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 120			
QY	571	GGGGCCAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 630			
DB	121	GGGGCCAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 180			
QY	631	CCCTGCCTGAGCGAGAGCTGAAGCTGCTCGAGTCCATCTTCCACCGCGGCGGAGCGC 690			
DB	181	CCCTGCCTGAGCGAGAGCTGAAGCTGCTCGAGTCCATCTTCCACCGCGGCGGAGCGC 240			
QY	691	TTCCGCATGCGAGCGCTGCTTGACAGCTGAGCTGCGAGTTCCTGCTGGTGGGCGC 750			
DB	241	TTCCGCATGCGAGCGCTGCTTGACAGCTGAGCTGCGAGTTCCTGCTGGTGGGCGC 300			
QY	751	GGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCT 810			
DB	301	GGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCT 360			
QY	811	GGGATCTGTCGCGCATCACTGCAACATCAGGAGTCATACCTGCTGTGCCCGCCATC 870			
DB	361	GGGATCTGTCGCGCATCACTGCAACATCAGGAGTCATACCTGCTGTGCCCGCCATC 420			
QY	871	TGGTCGGTGGAGTCTGTGACCCCTAACCTTGGCTGCTGTGGAGAGGCTGGTGACATA 930			
DB	421	TGGTCGGTGGAGTCTGTGACCCCTAACCTTGGCTGCTGTGGAGAGGCTGGTGACATA 480			
QY	931	AGAAGAAGGAATCTGCTATTGACAGCATCTGAAGAGATCATCTCCGACCTGTGTAAA 990			
DB	481	AGAAGAAGGAATCTGCTATTGACAGCATCTGAAGAGATCATCTCCGACCTGTGTAAA 540			

QY	991	CTCTATAACCTCCTCAGCATCCAGATGTGGAGATGCTGGATCAACCCCTTGCAGCAGAG 1050			
DB	541	CTCTATAACCTCCTCAGCATCCAGATGTGGAGATGCTGGATCAACCCCTTGCAGCAGAG 600			
QY	1051	CAGTCGACAGAGAACGCTGCTTCAGAGATGAAGATGAGGAGATGCTGAGGACACA 1110			
DB	601	CAGTCGACAGAGAACGCTGCTTCAGAGATGAAGATGAGGAGATGCTGAGGACACA 660			
QY	1111	GAAGACTTAGATCACTATGAAATGAAGAGAGAGAGCCAGCTGAGGCGCAAGAAATCTGAA 1170			
DB	661	GAAGACTTAGATCACTATGAAATGAAGAGAGAGAGCCAGCTGAGGCGCAAGAAATCTGAA 720			
QY	1171	GATGATGGCATGGAAGAAAACTTTGGCCATCCTAGAGAAAATTTAAAGAACACAGAG 1230			
DB	721	GATGATGGCATTGGAAAAAAGAACTTTGGCCATCCTAGAGAAAATTTAAAGAACACAGAG 780			
QY	1231	CAAGATTACTTAATGCTGAGTCTGCTGGCTCGGTGAGGCGCACTGACCGGCTGATGAAG 1290			
DB	781	CAAGATTACTTAATGCTGAGTCTGCTGGCTCGGTGAGGCGCACTGACCGGCTGATGAAG 840			
QY	1291	GAGCTCAGGGATATATACCGATCACAGAGTTTCAAGGGCGGAACTATGCGAGTCGAATC 1350			
DB	841	GAGCTCAGGGATATATACCGATCACAGAGTTTCAAGGGCGGAACTATGCGAGTCGAATC 900			
QY	1351	GTGAATGACAGTCTGTATGATGGAATGTCAAACCTCTCAAAGTTGACAGGACAGCGCT 1410			
DB	901	GTGAATGACAGTCTGTATGATGGAATGTCAAACCTCTCAAAGTTGACAGGACAGCGCT 960			
QY	1411	TTGCACAAACATCTCCAGATCTCTCAAGAGAAAGAGGAGCCGACATTCATCTTACCTAAC 1470			
DB	961	TTGCACAAACATCTCCAGATCTCTCAAGAGAAAGAGGAGCCGACATTCATCTTACCTAAC 1020			
QY	1471	TTTTTCCTTTAAAGATAACTTTTCCCTTTGACCCACCATTTTGTCTCAGGGTGTGTCTCCAGTC 1530			
DB	1021	TTTTTCCTTTAAAGATAACTTTTCCCTTTGACCCACCATTTTGTCTCAGGGTGTGTCTCCAGTC 1080			
QY	1531	CTCTCTGGAGGGATGTTCTGGGGGAGGGGCGCATCTGCATGGAATCTTCTACCAACAG 1590			
DB	1081	CTCTCTGGAGGGATGTTCTGGGGGAGGGGCGCATCTGCATGGAATCTTCTACCAACAG 1140			
QY	1591	GGCTGGAGAGTGCCTACTCTCATAGAGTCACTGATCATGATGAGTGCACACTGGTG 1650			
DB	1141	GGCTGGAGAGTGCCTACTCTCATAGAGTCACTGATCATGATGAGTGCACACTGGTG 1200			
QY	1651	AAGGGGAAAGACAGAGTGCAGTTTGGAGCCCAACAAATCTCAATACAGTCTGTACAAAGACA 1710			
DB	1201	AAGGGGAAAGACAGAGTGCAGTTTGGAGCCCAACAAATCTCAATACAGTCTGTACAAAGACA 1260			
QY	1711	CAGCAGTCTTACAGTCTTGGTGGTGCAGATCCAGAAAAAACGGCTGGTACACACCCCA 1770			
DB	1261	CAGCAGTCTTACAGTCTTGGTGGTGCAGATCCAGAAAAAACGGCTGGTACACACCCCA 1320			
QY	1771	AAAGAACAGCGCTAACCTTGAGTATCACCTTCTCCCTCCCGAGGACACACTGGACCA 1830			
DB	1321	AAAGAACAGCGCTAACCTTGAGTATCACCTTCTCCCTCCCGAGGACACACTGGACCA 1380			
QY	1831	ATTACCTTTGAATGCTGATTTGGATCTCAGCTGCTCTGTGGTTCCTCCCTCATTTT 1890			
DB	1381	ATTACCTTTGAATGCTGATTTGGATCTCAGCTGCTCTGTGGTTCCTCCCTCATTTT 1440			
QY	1891	TCCTGGAGTGTAGTCTTGCCTATTGACGAGCAATGATGGCTATTCTAAACGCTAAGGA 1950			
DB	1441	TCCTGGAGTGTAGTCTTGCCTATTGACGAGCAATGATGGCTATTCTAAACGCTAAGGA 1500			
QY	1951	AAAAAAACAACACAGAACTGTTTCAAGTACTCAAGTACTGACTTACAGACCAACCAACCA 2010			
DB	1501	AAAAAAACAACACAGAACTGTTTCAAGTACTCAAGTACTGACTTACAGACCAACCAACCA 1560			
QY	2011	CCTTGTCTGGAACCTTCTGACAGGCAATCTTATAAAGAAACTTTTCAGGCTCTCTTATA 2070			
DB	1561	CCTTGTCTGGAACCTTCTGACAGGCAATCTTATAAAGAAACTTTTCAGGCTCTCTTATA 1620			
QY	2071	TTGTGGAACACTAGCTGTGCTCCAGACTAGAGCCTCTCTTACCTATGATGGAATTTTA 2130			

Db 1621 TTGCTGGAACACTCAGCTGCTCCAGACTAGAGCTCTCTACCTATGATGATGATTTTA 1680
Qy 2131 ATTTATTTCTCTATTTTCATGACACAGCTGCTTTTGGTTACAGTGTATGATGATG 2190
Db 1681 ATTTATTTCTCTATTTTCATGACACAGCTGCTTTTGGTTACAGTGTATGATGATG 1740
Qy 2191 TATGAAAAAATGATCTTTTGGGAAAAAATACAGTTTGTATTTGAAAA 2242
Db 1741 TATGAAAAAATGATCTTTTGGGAAAAAATACAGTTTGTATTTGAAAA 1792

RESULT 2
AX017982

LOCUS AX017982 1880 bp DNA linear PAT 07-SEP-2000
DEFINITION ~~Sequence from Patent WO946375~~
ACCESSION AX017982
VERSION AX017982.1 GI:10042481
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarsky,C.
Human ~~nucleic acid~~ ~~sequences from~~ ~~postate tissue~~
Patent: ~~WO 946375~~ ~~6~~ ~~217~~ ~~10-SEP-1995~~
SCHMITT, ARMIN (DE); SPECHT, THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..1880
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 484 a 473 c 494 g 429 t
ORIGIN

Query Match 73.3%; Score 1652.4; DB 6; Length 1880;
Best Local Similarity 99.9%; Pred. No. 5.6e-195;
Matches 1664; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 587 CGCGGGGCGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 646
Db 1 CGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
Qy 647 AGCTGAAGCTGCTGAGTCCATCTT-CCACCGCGGCCACGAGCGCTTCCGATTTGCCAGC 705
Db 61 AGCTGAAGCTGCTGAGTCCATCTTCCACCGCGGCCACGAGCGCTTCCGATTTGCCAGC 120
Qy 706 GCCTGCTGGAGAGCTGAGCTGAGTGTCTGCTGGCTGGGCGGCGGAGGCGGCGGCGG 765
Db 121 GCCTGCTGGAGAGCTGAGCTGAGTGTCTGCTGGCTGGGCGGCGGAGGCGGCGGCGG 180
Qy 766 GGGCGCGCGCGGAGCCCATCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 825
Db 181 GGGCGCGCGCGGAGCCCATCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Qy 826 ATCCACTGCAACATCAGCGAGTCAATCCCTGCTGCGCCCGCCATCTGCTGGTGGAGTCT 885
Db 241 ATCCACTGCAACATCAGCGAGTCAATCCCTGCTGCGCCCGCCATCTGCTGGTGGAGTCT 300
Qy 886 GATGACCCCTAACTGGCTGCTGCTGTTGGAGGCTGTTGGACATAAAGAAAGGAATACT 945
Db 301 GATGACCCCTAACTGGCTGCTGCTGTTGGAGGCTGTTGGACATAAAGAAAGGAATACT 360
Qy 946 CTGCTATTGACGATCTGAAGAGATCATCTCCGACCTGTGTAACTCTATAACCTCCCT 1005
Db 361 CTGCTATTGACGATCTGAAGAGATCATCTCCGACCTGTGTAACTCTATAACCTCCCT 420
Qy 1006 CAGCATCCAGATGTGAGATGCTGGATCAACCCCTGGCAGCAGAGCAGTGCACACAGAA 1065
Db 1006 CAGCATCCAGATGTGAGATGCTGGATCAACCCCTGGCAGCAGAGCAGTGCACACAGAA 1065

Db 421 CAGCATCCAGATGTGAGATGCTGGATCAACCCCTGGCAGCAGAGCAGTGCACACAGAA 480
Qy 1066 GACGTGCTTTCAGAAAGATGAAGATGAGGAGATGCTGAGGACACAGAAAGACTTTAGATCAC 1125
Db 481 GACGTGCTTTCAGAAAGATGAAGATGAGGAGATGCTGAGGACACAGAAAGACTTTAGATCAC 540
Qy 1126 TATGAATGAAGAGGAGAGCCAGCTGAGGGCAAGAAATCTGAGAGATGATGGCATTTGA 1185
Db 541 TATGAATGAAGAGGAGAGCCAGCTGAGGGCAAGAAATCTGAGAGATGATGGCATTTGA 600
Qy 1186 AAAGAAACTTTGGCCATCTTAGAGAAATTTAAAGAAACAGAGGCGCAAGATTTACTTAAAT 1245
Db 601 AAAGAAACTTTGGCCATCTTAGAGAAATTTAAAGAAACAGAGGCGCAAGATTTACTTAAAT 660
Qy 1246 GGTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
Db 661 GGTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Qy 1306 TACCGATCAGAGATTTTCAAAGCGGGAACATATGACGTGCACTGCTGCTGCTGCTGCTGCT 1365
Db 721 TACCGATCAGAGATTTTCAAAGCGGGAACATATGACGTGCACTGCTGCTGCTGCTGCTGCT 780
Qy 1366 TATGATTTGGAATGTCAAACTCTCTCAAAAGTTGACCGAGCAGCGCTTTGCAACAGATCTC 1425
Db 781 TATGATTTGGAATGTCAAACTCTCTCAAAAGTTGACCGAGCAGCGCTTTGCAACAGATCTC 840
Qy 1426 CAGATCTCTCAAGAGAAAGAGCGGCGACTTCTTACTTACTTACTTACTTACTTACTTAAAGAT 1485
Db 841 CAGATCTCTCAAGAGAAAGAGCGGCGACTTCTTACTTACTTACTTACTTACTTACTTAAAGAT 900
Qy 1486 AACTTTCCCTTTGACCCACCATTTGTCAGGCTTGTGCTCCAGTCTCTCTCGAGGCTAT 1545
Db 901 AACTTTCCCTTTGACCCACCATTTGTCAGGCTTGTGCTCCAGTCTCTCTCGAGGCTAT 960
Qy 1546 GTTCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1605
Db 961 GTTCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
Qy 1606 TACTCCATAGAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
Db 1021 TACTCCATAGAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Qy 1666 GTGAGTTTGGAGCCCAACAATCTCAATACAGTCTGACAGAGCAGCAGCAGTCTCTACAG 1725
Db 1081 GTGAGTTTGGAGCCCAACAATCTCAATACAGTCTGACAGAGCAGCAGCAGTCTCTACAG 1140
Qy 1726 TCCTTGTGTGAGATCCACGAAAAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1785
Db 1141 TCCTTGTGTGAGATCCACGAAAAAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Qy 1786 CCCTGGAGTATCACCCCTTCCCTCCCGGAGCAGCAGTCTGAGCAGTCTGAGCAGTCTGAGCAG 1845
Db 1201 CCCTGGAGTATCACCCCTTCCCTCCCGGAGCAGCAGTCTGAGCAGTCTGAGCAGTCTGAGCAG 1260
Qy 1846 TGTATTTGGATCTCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1905
Db 1261 TGTATTTGGATCTCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Qy 1906 CTCTGCTTATTTGAGGACAATGATGCTTATTTAAAGCGCTTAAGAAAAAAGAAAAAAGAAAAA 1965
Db 1321 CTCTGCTTATTTGAGGACAATGATGCTTATTTAAAGCGCTTAAGAAAAAAGAAAAAAGAAAAA 1380
Qy 1966 GAACGTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTTCTGCTGGAACCCCT 2025
Db 1381 GAACGTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTTCTGCTGGAACCCCT 1440
Qy 2026 TGCTAGCAGGATCTTCTTATTAAGAAACTTTTCGAGCCTCTCTATATTTGCTGGAACCTAG 2085
Db 1441 TGCTAGCAGGATCTTCTTATTAAGAAACTTTTCGAGCCTCTCTATATTTGCTGGAACCTAG 1500
Qy 2086 CTGTGCTCCAGACTAGAGCCTCTTACCTATGCTATGATGATGATGATGATGATGATGATGATGAT 2145
Db 1501 CTGTGCTCCAGACTAGAGCCTCTTACCTATGCTATGATGATGATGATGATGATGATGATGATGAT 1560

D _b	1030	GCAGACTTCATCCTACTTAACATTCTCCTCTTAAAGATAAACATTCCCTTTTGACCACCCTTC	1089
Q _y	1510	GTCAAGGTTGTGTCTCCAGTCCCTCTCGAGGGTAGTGTCTTGGSCGSGAGSGGCATCTGC	1569
D _b	1090	GTCAAGGTTGTGTCTCCAGTCCCTCTCGAGGGTAGTGTCTTGGTGAGGTGCCACTCTGC	1149
Q _y	1570	ATGGAATCTTCTACCAAACAGGGCTGGAGCAGTGCCTACTCCATAGAGTCAGTGTGATCATG	1629
D _b	1150	ATGGAATCTTCTACCAAGCAGGGCTGGAGCAGTGCCTACTCCATAGAGTCCTGTGATCATG	1209
Q _y	1630	CAGATCAGTCCCACACTGTGTGAAGGGAAAGCACGAGTGCAGTTTTGSAGCCAACAATCT	1689
D _b	1210	CAGATCAGCCACGCTGTGTGAAGGGAAAGCACGGGTGCAGTTTGSAGCCAACAATCT	1269
Q _y	1690	CAATACAGTCTGACAGAGCACAGCAGTCTCTACAAGTCTTGGTGGAGTCCACAGAAAAA	1749
D _b	1270	CAGTATAGCTCGAGAGCACAGCAGTCTCTACAAGTCTTGGTGGAGTCCATGAAAAA	1329
Q _y	1750	AACGGCTGGTACACACCCCCAAAAGAGACGGCTAACCCCTGGAGTATCACCCCTCCCTCC	1809
D _b	1330	AACGGCTGGTACACACCCCCAAAAGAGATGGCTAGCCCTGSAGTCTCACCCCTCCT-CC	1388
Q _y	1810	TCCCCAGGCACCACTGGACCAATTACCTTTGAATGCTGTATTTGGATCTCACGCTGCCTC	1869
D _b	1389	TCCCCAGGCACCACTGGACCAATTACCTTTGAATGCTGTGTGGATCTCACGCTGCCTC	1448
Q _y	1870	TGTGGTTCCCTCCCTCATTTTTTCTGGACGTGATAGCTCTGCCTATTGTCAGGACAATGAT	1929
D _b	1449	TG-GTTTCCCCCTCATTTTTTCTGGACGTGATAGCTCTG-CTATTGTCAGGACAATGAT	1506
Q _y	1930	GGCTATTCTAAACGCTAAGGAAAAAACA-------ACACAGAATCTGTTTCAAGT	1979
D _b	1507	GGCTATTCTAAACGCTAAGGAAAAAACAACACCACACACACTGAACTGTCTCAAGT	1566
Q _y	1980	ACTCAAGCTGACTTACAGACCAACCAACCACTTGTGTGAAOCCCTTGTCTAGCAGGCAAT	2039
D _b	1567	ACTCAAGCTGACT- ----ACAGACCAACCACTTGTGTGAAOCCCTTGTCTAGCAGGCAAT	1621
Q _y	2040	CTTATAAAGAAACTTTCGAGCTCCTTATATGCTGGAACCTCAGCTGCTGCCAGACT	2099
D _b	1622	CTTAT-ANAGAAACTCTCAAGCCTCCTTATTTGCTGGAACCTCAGCTGCTGCCAGACT	1680
Q _y	2100	AGAGCCTCCTTACCTATGCTATGGATTTTTAAATTTATTTTCTCTATTTCATGTACACTG	2159
D _b	1681	AGAGCCTCCTTACCTATGCTATGGATTTTTAAATTTATTTTCTCTATTTCATGTACACTG	1740
Q _y	2160	CTTTTTTGGTTACAGTGTATGATGGATGTGTATGAAAAAAATGTATCTTTGGGAAAAACA	2219
D _b	1741	CTTTTTTGGTTTTCAGTGTATGATGGATGTGTATG-AAAAAATGTATCTTTGGGAAAAACA	1799
Q _y	2220	ATTACA-GTTTGTAAATTTGAAAA 2242	
D _b	1800	ATTACAGTTTGTAAATTTGAAAA 1823	

RESULT 5
 AX017833
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
 Pilarsky,C.
 Human nucleic acid sequences from prostate tissue
 Patent: ~~WO99/18373~~ ~~WO99/18373~~ ~~WO99/18373~~
 SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
 BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN


```
QY 1809 CTCGCCAGGACCACTGGACCAATACCTTTGAATGCTGTATTTGGATCTCACGCTGCT 1868
Db 901 CTCGCCAGGACCACTGGACCAATACCTTTGAATGCTGTATTTGGATCTCACGCTGCT 960
QY 1869 CTGTGGTTCCTCCCTCAATTTTCCTGGACGTGATAGCTGCTGCTATTTGCGAGGACAATGA 1928
Db 961 CTGTGGTTCCTCCCTCAATTTTCCTGGACGTGATAGCTGCTGCTATTTGCGAGGACAATGA 1020
QY 1929 TGGCTATTCTAAACGCTAAGGAAAAAACAACACACAGCAACTGTTTCAAGTACTCAAGAC 1988
Db 1021 TGGCTATTCTAAACGCTAAGGAAAAAACAACACAGCAACTGTTTCAAGTACTCAAGAC 1080
QY 1989 TGACTTACAGACCAACCAACCACTTGTCTGGAACCTTGCCTAGCAGGCAATCTTTATAAAA 2048
Db 1081 TGACTTACAGACCAACCAACCACTTGTCTGGAACCTTGCCTAGCAGGCAATCTTTATAAAA 1140
QY 2049 GAAACTTTCGAGCCTCCTTATATGCTGGAACCTCAGCTGCTGCCAGCTAGAGCCCTCC 2108
Db 1141 GAAACTTTCGAGCCTCCTTATATGCTGGAACCTCAGCTGCTGCCAGCTAGAGCCCTCC 1200
QY 2109 TTACTATGCTATGATGATTTTAAATTTATTTCTCTTATTTCTATGCTACACTGCTTTTTTG 2168
Db 1201 TTACTATGCTATGATTTTAAATTTATTTCTCTTATTTCTATGCTACACTGCTTTTTTG 1260
QY 2169 GTTACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2228
Db 1261 GTTACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 2229 TGTTAAATTTG 2238
Db 1321 TGTTAAATTTG 1330

RESULT 6
BD137243
LOCUS
DEFINITION Human nucleic acid sequence originating in prostatic tissue.
ACCESSION BD137243
VERSION BD137243.1 GI:23232188
KEYWORDS JP 2002505878-A/4.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1330)
AUTHORS Specht,T., Hinzmann,B., Schmitt,A., Pilarsky,C., Dahl,E. and
Rosenthal,A.
TITLE Human nucleic acid sequence originating in prostatic tissue
JOURNAL Patent: JP 2002505878-A 4 26-FEB-2002;
METAGEN GESELLSCHAFT FUER GENOM FORSCHUNG MBH
COMMENT OS Homo sapiens (human)
PN JP 2002505878-A/4
PD 26-FEB-2002
PF 09-MAR-1999 JP 2000535742
PR 10-MAR-1998 DE 198 11 194.0
PI THOMAS SPECHT, BERND HINZMANN, ARMIN SCHMITT, CHRISTIAN PILARSKY,
PI EDGAR DAHL,
PI ANDRE ROSENTHAL
PC C12N15/09, A61K38/00, A61K39/395, A61K48/00, A61P13/08,
PC A61P35/00,
PC C07K14/47, C07K16/18, C12N1/19, C12N1/21, C12N5/10, C12N15/02, PC
C12P21/02.
PC C12P21/08, G01N33/53, C12N15/00, A61K37/02, C12N5/00, C12N15/00 CC
Human nucleic acid sequence originating in prostatic tissue FH
key Location/Qualifiers
FT source 1. .1330
FT /organism="Homo sapiens (human)".
FEATURES
source Location/Qualifiers
1. .1330
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 394 a 304 c 303 g 329 t
```

```
ORIGIN
Query Match 59.0%; Score 1330; DB 6; Length 1330;
Best Local Similarity 100.0%; Pred. No. 4e-155;
Matches 1330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 909 CTTGGAGAGGCTGTGGACATAAAGAAAGGAATACCTCTGCTATTTGCGAGCATCTGAAAGAG 968
Db 1 CTTGGAGAGGCTGTGGACATAAAGAAAGGAATACCTCTGCTATTTGCGAGCATCTGAAAGAG 60
QY 969 GATCATCTCCGACCTGTGTAACTCTATAAACCCTCCCTCAGCATCCAGATGTGGAGATGCT 1028
Db 61 GATCATCTCCGACCTGTGTAACTCTATAAACCCTCCCTCAGCATCCAGATGTGGAGATGCT 120
QY 1029 GGATCAACCCCTGCCAGCAGCAGCAGTGCACACAGGAAGACGCTGCTTCAGAAAGATGAAGA 1088
Db 121 GGATCAACCCCTGCCAGCAGCAGCAGTGCACACAGGAAGACGCTGCTTCAGAAAGATGAAGA 180
QY 1089 TGAGGAGATGCTCAGGACACAGAGACTTAGATCACTATGAAATGAAGAGGAAGAGCC 1148
Db 181 TGAGGAGATGCTCAGGACACAGAGACTTAGATCACTATGAAATGAAGAGGAAGAGCC 240
QY 1149 AGCTGAGGGCAAGAAATCTGAAGATGATGGCAATTTGAAAGAAAGAAACTTTGGCCATCTCTAGA 1208
Db 241 AGCTGAGGGCAAGAAATCTGAAGATGATGGCAATTTGAAAGAAAGAAACTTTGGCCATCTCTAGA 300
QY 1209 GAAATTTAAAGAAACAGAGGCAAGATTTACTTAATGGTGCAGTGTCTGGCTCGGTGCA 1268
Db 301 GAAATTTAAAGAAACAGAGGCAAGATTTACTTAATGGTGCAGTGTCTGGCTCGGTGCA 360
QY 1269 GGCCACTGACCGGCTGATGAAGAGCTCAGGATATATACCGATCACAGATTTTCAAAAGG 1328
Db 361 GGCCACTGACCGGCTGATGAAGAGCTCAGGATATATACCGATCACAGATTTTCAAAAGG 420
QY 1329 CGGAAACTATGCACTGCAACTCTGTAATGACAGTCTCTATGATTTGGAATGTCAAACTCCT 1388
Db 421 CGGAAACTATGCACTGCAACTCTGTAATGACAGTCTCTATGATTTGGAATGTCAAACTCCT 480
QY 1389 CAAAGTTGACCAAGCAGGCGTTTGCAACAGATCTCCAGATCTCCAGATCTCCAGAAAGAGG 1448
Db 481 CAAAGTTGACCAAGCAGGCGTTTGCAACAGATCTCCAGATCTCCAGAAAGAGAGG 540
QY 1449 AGCCGACTTCAATCTTACTTAACTTTTCTTAAAGATAAATTTCCCTTTGACCCACCACAT 1508
Db 541 AGCCGACTTCAATCTTACTTAACTTTTCTTAAAGATAAATTTCCCTTTGACCCACCACAT 600
QY 1509 TGTGAGGTTGTGCTCTCCAGTCTCTCTGGAGGATGATGTTCTGGCGGAGGGGCGCATCTG 1568
Db 601 TGTGAGGTTGTGCTCTCCAGTCTCTCTGGAGGATGATGTTCTGGCGGAGGGGCGCATCTG 660
QY 1569 CATGGAACCTTCTACCAAAACAGGGCTGGAGCAGTGCCTACTTCCATAGAGTCAGTGCAT 1628
Db 661 CATGGAACCTTCTACCAAAACAGGGCTGGAGCAGTGCCTACTTCCATAGAGTCAGTGCAT 720
QY 1629 GCAGATCAGTCCCACTGTTGAGGGGAAAGCAGCAGTGCAGTTTGGAGCCCAACAATC 1688
Db 721 GCAGATCAGTCCCACTGTTGAGGGGAAAGCAGCAGTGCAGTTTGGAGCCCAACAATC 780
QY 1689 TCAATACAGTCTGACAAAGACACAGCAGTCTCTACAACTCTCTGGTGAGATCCACGAAAA 1748
Db 781 TCAATACAGTCTGACAAAGACACAGCAGTCTCTACAACTCTCTGGTGAGATCCACGAAAA 840
QY 1749 AAACGGCTGTACACACCCCAAAAGAGCGGCTAAACCTGGAGTATCACCCCTTCCTCC 1808
Db 841 AAACGGCTGTACACACCCCAAAAGAGCGGCTAAACCTGGAGTATCACCCCTTCCTCC 900
QY 1809 CTCCCCAGGACCACTCGACCAATTTACTTTGAATGCTGTATTTGGATCTCAGCTGCCT 1868
Db 901 CTCCCCAGGACCACTCGACCAATTTACTTTGAATGCTGTATTTGGATCTCAGCTGCCT 960
QY 1869 CTGTGGTTCCTCCCTCAATTTTCCTGGACGTGATAGCTGCTGCTATTTGCGAGGACAATGA 1928
Db 961 CTGTGGTTCCTCCCTCAATTTTCCTGGACGTGATAGCTGCTGCTATTTGCGAGGACAATGA 1020
```



```

QY 1929 TGGCTATTCTAAAGCGTAAGGAAAAAACAACACAGAACTGTTCAAGTACTCAAGAC 1988
Db 1021 TGGCTATTCTAAAGCGTAAGGAAAAAACAACACAGAACTGTTCAAGTACTCAAGAC 1080
QY 1989 TGACTTACAGACCAACCAACCACTTGTGGAACCCCTTGCTAGCAGGCAATTCCTATATAA 2048
Db 1081 TGACTTACAGACCAACCAACCACTTGTGGAACCCCTTGCTAGCAGGCAATTCCTATATAA 1140
QY 2049 GAAACTTTTCGAGCGCTCTTATATTGCTGGAACCTCAGCTGTGCTCCAGACTAGAGCCTCC 2108
Db 1141 GAACTTTTCGAGCGCTCTTATATTGCTGGAACCTCAGCTGTGCTCCAGACTAGAGCCTCC 1200
QY 2109 TTACTATGCTATGAGATTTTAAATTAATTTCTTATTTCATGATACACTGCTTTTTTGG 2168
Db 1201 TTACTATGCTATGAGATTTTAAATTAATTTCTTATTTCATGATACACTGCTTTTTTGG 1260
QY 2169 GTTACAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2228
Db 1261 GTTACAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 2229 TGTAAATTTG 2238
Db 1321 TGTAAATTTG 1330

```

```

RESULT 7
BC000848
LOCUS BC000848 1162 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3458173, mRNA, partial cds.
ACCESSION BC000848
VERSION BC000848.1 GI:12654076
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1162)
Strausberg, R.
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. W.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAP Plate: 4 Row: 1 Column: 6

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6688150.

```

FEATURES
source
1..1162
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:31214"
/db_xref="Taxon:9606"
/clone="IMAGE:3458173"
/tissue_type="Cervix, carcinoma"
/clone_lib="NIH_MGC_12"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

```

CDS

```

<1..694
/codon_start=2
/product="Unknown (protein for IMAGE:3458173)"
/protein_id="AA00848.1"
/db_xref="GI:12654077"
/translaton="EMPEDEDLDHYEMKEEPEAEKKSEDDGIGKENTAILKIKKN
ORDYLNQAVSGVQATRLMKELDIYRSQSPKGNAYVELYDNLVSLYDNNVLLKVD
QDSALHNDLQILKEKEGADPILNLFQKDFPFPFVSVPLVSGVYVGGGATCM
ELUTKQWSSAYSIESVIMQISATLVKGRVQFGKANSQYSLTRAQSQSYKSLVQIHE
KNHWYTPPEEDG"
BASE COUNT 355 a 263 c 254 g 290 t
ORIGIN

Query Match 51.5%; Score 1160.4; DB 9; Length 1162;
Best Local Similarity 99.9%; Pred. No. 3.7e-134;
Matches 1161; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1092 GGAGATGCTGAGGACACAGAACTTAGATCATCTAGTAATGAAGAGGAAGAGGCGC 1151
Db 1 GGAGATGCTGAGGACACAGAACTTAGATCATCTAGTAATGAAGAGGAAGAGGCGC 60

QY 1152 TGAGGCAAGAAATCTGAAGATGATGCTTGAAGAGAACTTGCCCATCTCAGAGAA 1211
Db 61 TGAGGCAAGAAATCTGAAGATGATGCTTGAAGAGAACTTGCCCATCTCAGAGAA 120

QY 1212 AATTAAAGAACACAGAGGCAAGATTACTTAAATGGTGCAGTGTCTGGCTCGGTGCAGGC 1271
Db 121 AATTAAAGAACACAGAGGCAAGATTACTTAAATGGTGCAGTGTCTGGCTCGGTGCAGGC 180

QY 1272 CACTGACCGGCTGATGAAGAGGCTCAGGATATATACCGATCACAGATTTCAAGGCGG 1331
Db 181 CACTGACCGGCTGATGAAGAGGCTCAGGATATATACCGATCACAGATTTCAAGGCGG 240

QY 1332 AACTATGCACTCGAATCTGTAATGACAGCTCTGTATGATGGAATGCAAACTCTCA 1391
Db 241 AACTATGCACTCGAATCTGTAATGACAGCTCTGTATGATGGAATGCAAACTCTCA 300

QY 1392 AGTTGACAGGACAGCGCTTTGCACACGATCTCCAGATCTCCAGATCTCCAGATCTCCAG 1451
Db 301 AGTTGACAGGACAGCGCTTTGCACACGATCTCCAGATCTCCAGATCTCCAGATCTCCAG 360

QY 1452 CGACTTCATCTACTTAACCTTTCCCTTAAAGATACTTCCCTTTGACCCACCATTTGT 1511
Db 361 CGACTTCATCTACTTAACCTTTCCCTTAAAGATACTTCCCTTTGACCCACCATTTGT 420

QY 1512 CAGGTTGTCTCTCCAGTCTCTCTGGAGGTATGTTCTGGGGAGGGGCGCATCTGCAT 1571
Db 421 CAGGTTGTCTCTCCAGTCTCTCTGGAGGTATGTTCTGGGGAGGGGCGCATCTGCAT 480

QY 1572 GGAATTTCTCACCACAGAGGCTGGAGCAGTGCCTACTCCATAGAGTCAATGATCA 1631
Db 481 GGAATTTCTCACCACAGAGGCTGGAGCAGTGCCTACTCCATAGAGTCAATGATCA 540

QY 1632 GATCAGTGCCACACTGTTGAAGGGGAAAGCAGGTCGAGTCTGGAGCCCAACAATCTCA 1691
Db 541 GATCAGTGCCACACTGTTGAAGGGGAAAGCAGGTCGAGTCTGGAGCCCAACAATCTCA 600

QY 1692 ATACAGTCTCACAAGAGCAGCAGTCTCTAAGTCTCTTGGTGCAGATCCACGAAAAAAA 1751
Db 601 ATACAGTCTCACAAGAGCAGCAGTCTCTAAGTCTCTTGGTGCAGATCCACGAAAAAAA 660

QY 1752 CGGCTGTACACACCCCCCAAGAACGCGCTAACCCCTGGAGTATCACCCCTTCCCTCCTC 1811
Db 661 CGGCTGTACACACCCCCCAAGAACGCGCTAACCCCTGGAGTATCACCCCTTCCCTCCTC 720

QY 1812 CCAGGACCACTGGACCAATTTACCTTTGAATGCTGTATTTGGATCTCAGCTGCTCTG 1871
Db 721 CCAGGACCACTGGACCAATTTACCTTTGAATGCTGTATTTGGATCTCAGCTGCTCTG 780

QY 1872 TGGTTCCTCCCTCATTTTCCCTGGAGGTGATAGTCTGCTATTTGCAGCAATGATGG 1931
Db 781 TGGTTCCTCCCTCATTTTCCCTGGAGGTGATAGTCTGCTATTTGCAGCAATGATGG 840

```


[illegible]

JOURNAL

Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Marcello Bento Soares, Ph.D.

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

Contact: amadansystemsbiology.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 66 Row: a Column: 14

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

FEATURES

source

Location/Qualifiers

1. .1173

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:105437"

/tissue_type="Mammary gland"

/clone_lib="Soares_mammary_gland_NbMMG"

/lab_host="DH10B"

/note="Vector: pT73D-Pac"

<1. .697

/codon_start=2

/product="RIKEN cDNA 2310012M18 gene"

/protein_id="AAH30044.2"

/db_xref="GI:28175091"

/db_xref="LocusID:70093"

/translation="EEMPEPTDLDHYEMKEEPAEGKSEDDGIGKENLAILEKIKK

NRQDYLNAGVRSQVQATDRMLKELDIYRSQFKGNYAVELVNDLSYDWNVKKLV

DQSDALNDLIQLEKEGADFIILNFSFKDNPFPFVRVSVPLSGVYVLSGGAIC

MEILLTKGWSAYSIESVIMQISATLVKGRVQFGANKSQYSILTRAQYSKSLVQIH

EKNWTPPPEDG"

BASE COUNT 347 a 275 c 266 g 285 t

ORIGIN

Query Match 44.3%; Score 998; DB 10; Length 1173;

Best Local Similarity 94.1%; Pred. No. 4.3e-114;

Matches 1107; Conservative 0; Mismatches 50; Indels 19; Gaps 6;

QY 1089 TGAGGAGATGCTGAGGACACAGAGACTTAGATCACTATGAATGAAGAGGAGGCC 1148

Db 1 TGAAGAGATGCTGAGGATACAGAGACCTAGATCACTATGAATGAAGAGGAGGCC 60

QY 1149 AGCTAGGCGCAAGAAATCTGAAGATGATGGCAATTTGGAAGAAACATTTGCCATCCTAGA 1208

Db 61 AGCGGAGGCGAAGAGTCTGAAGATGATGGCATCGGAAGAGAACCTGGCCATCCTAGA 120

QY 1209 GAAATTAAGAAACACAGAGCGCAAGATTACTTAATGGTGCAGTGTCTGCTCGGTGCA 1268

Db 121 GAAATTAAGAAACACAGAGCGCAAGATTACTTAATGGTGCAGTGTCTGCTCGGTGCA 180

QY 1269 GGCCACTGACCGGTGATGAAGAGCTCAGGATATATACCGATCACAGAGCTTCAAGG 1328

Db 181 GGCCACTGACCGGTGATGAAGAGCTCAGGATATATACCGATCACAGAGCTTCAAGG 240

QY 1329 CGGAAATATGCACTGCAACTCGTGAATGACAGTCTGTATGATTTGGAATGCAACTCCT 1388

Db 241 CGGAAATATGCACTGCAACTCGTGAATGACAGTCTGTATGATTTGGAATGCAACTCCT 300

QY 1389 CAAAGTTGACGACAGCGCTTTGCAACAGATCTCCAGATCTCCAGATCTCCAGAGAGG 1448

|||||

301 CAAAGTTGACGACGATAGCCCTTTGCACATGATCTTCAGATCCTGAAGAGAGGAGG 360

QY 1449 AGCGGACTTCATCTCTACTTAACCTTTTCTTTAAAGATAAATTTCCCTTTGACCCACCAT 1508

Db 361 AGCAGACTTCATCTCTACTTAACCTTTTCTTTAAAGATAAATTTCCCTTTGACCCACCAT 420

QY 1509 TGTCAAGGTTGTCTCTCCAGTCTCTCTGGAGGATGTTCTGGGGGAGGGGCCATCTG 1568

Db 421 CGTCAGGTTGTGTCTCCAGTCTCTCTGGAGGATGTTCTGGGGGAGGGGCCATCTG 480

QY 1569 CATGGAATCTCTCACCACAAACAGGCTGGAGCAGTGCCTACTCCATAGAGTCACTGATCAT 1628

Db 481 CATGGAATCTCTCACCACAAACAGGCTGGAGCAGTGCCTACTCCATAGAGTCCGCTGATCAT 540

QY 1629 GCAGATCAGTGCACACTGTGTGAAGGGGAAAGCAGAGTGCAGTGTGGAGCCCAACAATC 1688

Db 541 GCAGATCAGGCGCCAGCTGTGTGAAGGGGAAAGCAGGCTGCAGTGTGGAGCCCAACAATC 600

QY 1689 TCAATACAGTCTGCACAGACACAGAGTCTCTCAAGTCTCTGGTGCAGATCCACGAAAA 1748

Db 601 TCAATACAGTCTGCACAGACACAGAGTCTCTCAAGTCTCTGGTGCAGATCCACGAAAA 660

QY 1749 AAACGGCTGTACACACCCCAAGAGAGCGCTAACCTGAGATATCACCTTCTCTCC 1808

Db 661 AAACGGCTGTACACACCCCAAGAGAGCGCTAACCTGAGATATCACCTTCTCTCC-C 719

QY 1809 CTCCCCAGGACCACTGGACCAATTAACCTTTGAATGCTGTATTGGATCTCAGCGTGCCT 1868

Db 720 CTCCCCAGGACCACTGGACCAATTAACCTTTGAATGCTGTATTGGATCTCAGCGTGCCT 779

QY 1869 CTGTGGTCTCTCCCTCATTTTCTGGAGCTGATAGCTCTGCTATTTGAGGACAAATGA 1928

Db 780 CTG-GTTTCCCTCCCTCATTTTCTGGAGCTGATAGCTCTGCTATTTGAGGACAAATGA 838

QY 1929 TGGCTATTCTAAAGCGCTAAGGAAAAAACA-----ACACAGAACTGTTCACAG 1978

Db 839 TGGCTATTCTAAGCGCTAAGGAAAAAACA-----ACACAGAACTGTTCACAG 898

QY 1979 TACTCAAGTCTGATACAGCAACCAACCACTTGTCTGGAACCTTGTAGCAGCAT 2038

Db 899 TACTCAAGTCTGAT-----ACAGACCAACCACTTGTCTGGAACCTTGTAGCAGCAT 953

QY 2039 TCTTATAAAGAAACCTTCAGCGCTCTTATATGCTGGAAACTCAGCTGTCTCCAGAC 2098

Db 954 TCTTAT-AAAGAAACTCTCAAGCTCTTATATGCTGGAAACTCAGCTGTCTCCAGAC 1012

QY 2099 TAGAGCTCTTACCTATGCTATGGATTTTAAATTTATTTCTCTTATTTTCTGATCAT 2158

Db 1013 TAGAGCTCTTACCTATGCTATGGATTTTAAATTTATTTCTCTTATTTTCTGATCAT 1072

QY 2159 GCTTTTGTGTTTACAGTGTATGATGGATGTATGAAAAAATGATATCTTTTGGGAAAAAC 2218

Db 1073 GCTTTTGTGTTTACAGTGTATGATGGATGTATGAAAAAATGATATCTTTTGGGAAAAAC 1131

QY 2219 AATTACAGTTTGTAAATTTGAAAAAATAAAAAA 2254

Db 1132 AATTACAGTTTGTAAATTTGAAAAAATAAAAAA 1167

RESULT 11

BC009286

LOCUS

DEFINITION

complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BC009286 Homo sapiens, NICE-5 protein, clone MGC:14087 IMAGE:3927447, mRNA,

complete cds.

ACCESSION BC009286 GI:14424527

VERSION BC009286.1

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 992)

STRAUSBERG, R.

TITLE
JOURNAL

REMARK
COMMENT

Direct Submission
Submitted (12-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Ruben Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IMAI Plate: 20 Row: b Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6688150.

FEATURES
source

1..992
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:55585"
/db_xref="taxon:9606"
/clone="MGC:14087 IMAGE:3927447"
/tissue_type="Ovary adenocarcinoma"
/clone_lib="NIH_MGC_9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
292..507
/codon_start=1
/product="NICE-5 protein"
/protein_id="AAH09286.1"
/db_xref="GI:14424528"
/translation="MELTIKQWSSAYSIESVIMQISLIVKGRVQFGANKSOYSL
TRAQSYKSLVQIHEKNGWYTPPKEDG"
BASE COUNT 302 a 232 c 203 g 255 t
ORIGIN

Query Match 43.0%; Score 969.4; DB 9; Length 992;
Best Local Similarity 99.9%; Pred. No. 1.5e-110;
Matches 970; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1284 GATGAGGAGCTCAGGATATATACCGATACAGATTTCAAGCGGAACTATGAGT 1343
DB 6 GAGGAGGAGCTCAGGATATATACCGATACAGATTTCAAGCGGAACTATGAGT 65
QY 1344 CGAACTCGTGAATGACAGTCTGTATGATTGGAATGTCAAACTCTCAAGTTGACCAAGGA 1403
DB 66 CGAACTCGTGAATGACAGTCTGTATGATTGGAATGTCAAACTCTCAAGTTGACCAAGGA 125
QY 1404 CAGCGCTTGCACACGATCTCCAGATCTCTCAAGAGAGAAAGAGCGGACTTCATTCT 1463
DB 126 CAGCGCTTGCACACGATCTCCAGATCTCTCAAGAGAGAAAGAGCGGACTTCATTCT 185
QY 1464 ACTTAACCTTTCCTTAAAGATAACTTCCCTTTGACCCACCATTTGTCAGGGTTGTGTC 1523
DB 186 ACTTAACCTTTCCTTAAAGATAACTTTCCTTTGACCCACCATTTGTCAGGGTTGTGTC 245
QY 1524 TCCAGTCTCTCTGAGGATATGTTCTGGCGGAGGGGCCATCTGCATGGAATCTTCAC 1583
DB 246 TCCAGTCTCTCTGAGGATATGTTCTGGCGGAGGGGCCATCTGCATGGAATCTTCAC 305

QY 1584 CAAACAGGGCTGGAGCAGTCCCTACTCCATAGAGTCAAGTATCATGAGATCATGCGCCAC 1643
DB 306 CAAACAGGGCTGGAGCAGTCCCTACTCCATAGAGTCAAGTATCATGAGATCATGCGCCAC 365
QY 1644 ACTGGTCAAGGGGAAAGCAGCAGTGGAGTGGAGCAACAAATCTCAATACAGTCTGAC 1703
DB 366 ACTGGTGAAGGGGAAAGCAGCAGTGGAGTGGAGCAACAAATCTCAATACAGTCTGAC 425
QY 1704 AAGACACACAGTCTCTACAAAGTCCCTGGTGGAGTCCAGAGTCCAGAGTCCAGAGTCCAC 1763
DB 426 AAGACACACAGTCTCTACAAAGTCCCTGGTGGAGTCCAGAGTCCAGAGTCCAGAGTCCAC 485
QY 1764 ACCCCCAAAAGAGAGCGCTAACCCCTGGAGTATCACCCCTTCCCTCCCTCCAGCAGCAC 1823
DB 486 ACCCCCAAAAGAGAGCGCTAACCCCTGGAGTATCACCCCTTCCCTCCCTCCAGCAGCAC 545
QY 1824 TGGACCAATTAACCTTTGAATGCTGATTTGGATCTCAGCTGCTCTGTGTTCCCTCC 1883
DB 546 TGGACCAATTAACCTTTGAATGCTGATTTGGATCTCAGCTGCTCTGTGTTCCCTCC 605
QY 1884 TCATTTTCTCGACGCTGATAGCTCTGCTTATTCAGGACAATGATGGCTATTCTAAACG 1943
DB 606 TCATTTTCTCGACGCTGATAGCTCTGCTTATTCAGGACAATGATGGCTATTCTAAACG 665
QY 1944 CTAAGGAAAAAACAACACAGACTGTTTCAAGTACTCAAGTACTCAAGTACTCAAGTACTCA 2003
DB 666 CTAAGGAAAAAACAACACAGACTGTTTCAAGTACTCAAGTACTCAAGTACTCAAGTACTCA 725
QY 2004 CCACACACCTTGGTGGACCTTGGTGGAGGATCTTATTAAGAGAACTTTTCGAGCCT 2063
DB 726 CCACACACCTTGGTGGACCTTGGTGGAGGATCTTATTAAGAGAACTTTTCGAGCCT 785
QY 2064 CCTATATTGCTGGAACCTCAGCTGCTCCAGACTAGAGCCTCTTACCTATGCTATGG 2123
DB 786 CCTATATTGCTGGAACCTCAGCTGCTCCAGACTAGAGCCTCTTACCTATGCTATGG 845
QY 2124 ATTTTAAATTTATTTCTCTTATTCATGATGATGCTGCTTTTGGTGTACAGTATGAT 2183
DB 846 ATTTTAAATTTATTTCTCTTATTCATGATGATGCTGCTTTTGGTGTACAGTATGAT 905
QY 2184 GGATGCTATGAAAAAATCTATCTTTGGGAAAAAATGATCTTTGGGAAAAAATGATCTTTG 2243
DB 906 GGATGCTATGAAAAAATGATCTTTGGGAAAAAATGATCTTTGGGAAAAAATGATCTTTG 965
QY 2244 AAAAAAATAA 2254
DB 966 AAAAAAATAA 976

RESULT 12
HSA243666
LOCUS HSA243666 885 bp mRNA linear PRI 18-MAR-2001
DEFINITION Homo sapiens mRNA for NICE-5 protein.
ACCESSION AJ243666
VERSION AJ243666.1 GI:6688150
KEYWORDS NICE-5 gene; NICE-5 protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
Marenholz, I., Zirma, M., Fischer, D.F., Backendorf, C., Ziegler, A. and
Mischke, D.
Identification of human epidermal differentiation complex
(EDC)-encoded genes by subtractive hybridization of entire YACS to
a gridded keratinocyte cDNA library
Genome Res. 11 (3), 341-355 (2001)
JOURNAL MEDLINE 21154910
PUBMED 11230159
REFERENCE 2 (bases 1 to 885)
Marenholz, I.
Direct Submission

JOURNAL Submitted (30-JUN-1999) Marenholz I., Universitaetsklinikum
Charite, Humboldt-Universitaet Berlin, Institut fuer Immunogenetik,
Spandauer Damm 130, 14050 Berlin, GERMANY

FEATURES
source
1..885
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/map=1q21
/clone="115918"
/cell_type="primary keratinocytes"
/clone_lib="KER 1/2"
1..885
/gene="NICE-5"
212..427
/gene="NICE-5"
/codon_start=1
/product="NICE-5 protein"
/protein_id="CA65097.1"
/db_xref="GI:6688151"
/db_xref="SPTREMBL:Q9UGL6"
/translation="MELTRKQSWSSIESVIMQISATLVKRGARVQFGANKSQYSL
TRAQSYKSLVQIHEKNGWYTPPKEDG"
881

polya_site
/gene="NICE-5"
247 a 215 c 180 g 243 t

BASE COUNT 247 a 215 c 180 g 243 t

ORIGIN

Query Match 39.1%; Score 880.8; DB 9; Length 885;
Best Local Similarity 99.8%; Pred. No. 1.4e-99;
Matches 882; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1360 AGCTGTATGATGGAATGCTCAAACTCCTCAAAGTTGACGAGCAGCGCTTTGCACAAC 1419
DB 2 AGCTGTATGATGGAATGCTCAAACTCCTCAAAGTTGACGAGCAGCGCTTTGCACAAC 61

QY 1420 GATCTCCAGATCCTCAAGAGAAAGAGGAGCGACTTCATCTCTACTTAACCTTTCCCTTT 1479
DB 62 GATCTCCAGATCCTCAAGAGAAAGAGGAGCGCGACTTCATCTCTACTTAACCTTTCCCTTT 121

QY 1480 AAGATAACTTTCCCTTTGACCCACCATTTGTGAGGGTTGTCTCCAGTCTCTCTGGA 1539
DB 122 AAGATAACTTTCCCTTTGACCCACCATTTGTGAGGGTTGTCTCCAGTCTCTCTGGA 181

QY 1540 GGGTATGTTCTGGGGGAGGGGCCATCTGCATGGAATCTCACCACACAGGGCTGGAGC 1599
DB 182 GGGTATGTTCTGGGGGAGGGGCCATCTGCATGGAATCTCACCACACAGGGCTGGAGC 241

QY 1600 AGTGCTACTCCATAGATGTCAGTGATCATGACAGATCAGTGCCACACTGGTGAAGGGAAA 1659
DB 242 AGTGCTACTCCATAGATGTCAGTGATCATGACAGATCAGTGCCACACTGGTGAAGGGAAA 301

QY 1660 GCACGAGTGCAGTTTGGAGCCCAACAAATCTCAATACAGTCTGACAGAGCAGCAGTCC 1719
DB 302 GCACGAGTGCAGTTTGGAGCCCAACAAATCTCAATACAGTCTGACAGAGCAGCAGTCC 361

QY 1720 TACAAGTCTTGGTGCAGATCCACGAAAGAAAGAGCGTGGTACACACCCCAAGAGAAC 1779
DB 362 TACAAGTCTTGGTGCAGATCCACGAAAGAAAGAGCGTGGTACACACCCCAAGAGAAC 421

QY 1780 GGCTAACCTCGAGTATCACCTTCCCTCCCTCCAGGACACCTGGACCAATACCTTT 1839
DB 422 GGCTAACCTCGAGTATCACCTTCCCTCCCTCCAGGACACCTGGACCAATACCTTT 481

QY 1840 GAATGCTGTATTTGGATCTCAGCTGCGCTCTGTGGTTCCCTCCCTCAATTTTCCCTGGAGC 1899
DB 482 GAATGCTGTATTTGGATCTCAGCTGCGCTCTGTGGTTCCCTCCCTCAATTTTCCCTGGAGC 541

QY 1900 TGATAGCTCTGCTTATTCAGGACAAATGATGCTATTTCTAAACGCTAAGGAAAAAACA 1959
DB 542 TGATAGCTCTGCTTATTCAGGACAAATGATGCTATTTCTAAACGCTAAGGAAAAAACA 601

QY 1960 AACACAGAACTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTTGTCTGG 2019
DB 602 AACACAGAACTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTTGTCTGG 661

QY 2020 AACCCCTGCTAGCAGGCACTTCTTATAAAGAAACTTTTCGAGCCTCCTTATATTTGCTGGA 2079
DB 662 AACCCCTGCTAGCAGGCACTTCTTATAAAGAAACTTTTCGAGCCTCCTTATATTTGCTGGA 721

QY 2080 ACTCAGCTGTCTCCAGACTAGAGCCTCCTTACCTATGCTATGCTATGGAATTTTATTTATTTT 2139
DB 722 ACTCAGCTGTCTCCAGACTAGAGCCTCCTTACCTATGCTATGCTATGGAATTTTATTTATTTT 781

QY 2140 CTCCTATTTTCATGACACTGCTTTTGGTTCAGAGTGTATGATGATGCTATCAAAAA 2199
DB 782 CTCCTATTTTCATGACACTGCTTTTGGTTCAGAGTGTATGATGATGCTATGATGAAAA 841

QY 2200 AATGTATCTTTGGGAAAAACAATTACAGTTTCTTAATTTGAAAAA 2243
DB 842 AATGTATCTTTGGGAAAAACAATTACAGTTTCTTAATTTGAAAAA 885

RESULT 13
AF116721 standard; RNA; HUM; 808 BP.
XX AF116721;
XX AF116721.1
XX AF116721.1

DT 24-MAY-2000 (Rel. 63, Created)
DT 09-MAY-2001 (Rel. 67, Last updated, Version 2)
XX
DE Homo sapiens PRO3094 mRNA, complete cds.
XX
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
XX
RN [1]
RP 1-808
RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,
RA Liu M., He F.;
RT "Functional prediction of the coding sequences of 121 new genes deduced by
RT analysis of cDNA clones from human fetal liver";
RL Unpublished.
XX
RN [2]
RP 1-808
RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,
RA Liu M., He F.;
RT Submitted (24-DEC-1998) to the EMBL/GenBank/DBJ databases.
RL Department of Experimental Hematology, Institute of Radiation Medicine,
RL Beijing Taiping Road 27, Beijing 100850, P. R. China
XX
DR SPTREMBL; Q9UGL6; Q9UGL6.
XX
FH Key Location/Qualifiers
FT source 1..808
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="FLC1398"
FT /tissue_type="liver"
FT /dev_stage="fetal"
FT 121..336
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9UGL6"
FT /note="predicted protein of HQ3094"
FT /product="PRO3094"
FT /protein_id="AAF71141.1"
FT FT


```
FT /translation="MELLFKQGHSSAYSIESVIMQISATLVKARVQFGANKSQYSLT
FT RAQSQYKSLVQIIEKNGWYTPPKEDG"
XX
SQ Sequence 808 BP; 231 A; 196 C; 159 G; 222 T; 0 other;

Query Match 35.2%; Score 794; DB 17; Length 808;
Best Local Similarity 99.9%; Pred. No. 7.4e-89;
Matches 805; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1450 GCCGACTTCATCTTACTTAACTTTCTCTTAAAGATACATTTCCCTTTGACCCACCATTT 1509
DB 1 GCCGACTTCATCTTACTTAACTTTCTCTTAAAGATACATTTCCCTTTGACCCACCATTT 60

QY 1510 GTCAGGTTGTCTCCAGTCCTCTCTGGAGGTATGTCTTGGCGGAGGGCCATCTGC 1569
DB 61 GTCAGGTTGTCTCCAGTCCTCTCTGGAGGTATGTCTTGGCGGAGGGCCATCTGC 120

QY 1570 ATGGAATCTTCACCAACAGGGCTGGAGAGTGCCTTACTCCATAGATCAGTGCATG 1629
DB 121 ATGGAATCTTCACCAACAGGGCTGGAGAGTGCCTTACTCCATAGATCAGTGCATG 180

QY 1630 CAGATCAGTGCACACTGTGTAAGGGGAAAGCAGAGTGCAGTTTGGAGCCCAACAATCT 1689
DB 181 CAGATCAGTGCACACTGTGTAAGGGGAAAGCAGAGTGCAGTTTGGAGCCCAACAATCT 240

QY 1690 CAATACAGTCTGACAGACAGCAGTCTCTCAAGTCTTGTGTCAGATCCAGCAAAA 1749
DB 241 CAATACAGTCTGACAGACAGCAGTCTCTCAAGTCTTGTGTCAGATCCAGCAAAA 300

QY 1750 AACGGTGTGACACACCCCAAGAGACGGCTTAACCTGGAGATACACCTTCCCTCC 1809
DB 301 AACGGTGTGACACACCCCAAGAGACGGCTTAACCTGGAGATACACCTTCCCTCC 360

QY 1810 TCCCGAGCACCCTGGACCAATACCTTTGAATGCTGATTTGAGTCTACGCTGCCTC 1869
DB 361 TCCCGAGCACCCTGGACCAATACCTTTGAATGCTGATTTGAGTCTACGCTGCCTC 420

QY 1870 TGTGTTCCCTCCCTCATTTTCTTGGAGTGTATGCTCTGCTTATGAGCAACAATGAT 1929
DB 421 TGTGTTCCCTCCCTCATTTTCTTGGAGTGTATGCTCTGCTTATGAGCAACAATGAT 480

QY 1930 GGCTATCTTAACGCTAAGG-AAAGAAACAAACAGACAGTCTTCAAGTACTCAAGAC 1988
DB 481 GGCTATCTTAACGCTAAGGAAAGAAACAAACAGACAGTCTTCAAGTACTCAAGAC 540

QY 1989 TGACTTACAGACCAACCAACCTTGTGGAACCTTGTAGCAGGATCTTATATAAA 2048
DB 541 TGACTTACAGACCAACCAACCTTGTGGAACCTTGTAGCAGGATCTTATATAAA 600

QY 2049 GAAACTTTCGAGCTCTTATATTGCTGGAACTCAGCTGTGCTCCAGACTAGAGCTCC 2108
DB 601 GAAACTTTCGAGCTCTTATATTGCTGGAACTCAGCTGTGCTCCAGACTAGAGCTCC 660

QY 2109 TTACTATGCTATGATTTTAAATTTATTTCTTCTTATTTCTATGATGACTCTTTTGG 2168
DB 661 TTACTATGCTATGATTTTAAATTTATTTCTTCTTATTTCTATGATGACTCTTTTGG 720

QY 2169 GTTACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2228
DB 721 GTTACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780

QY 2229 TGTAAATTTGAAAAAAGAAAAA 2254
DB 781 TGTAAATTTGAAAAAAGAAAAA 806

RESULT 14
BC015316
LOCUS Homo sapiens, NICE-5 protein, clone MGC:21081 IMAGE:4151953, mRNA,
DEFINITION complete cds.
ACCESSION BC015316
VERSION BC015316.1 GI:15929786
```

```
KEYWORDS MGC. Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 772)
JOURNAL Strausberg R.
Direct Submission
Submitted (01-OCT-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcaps-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 20 Row: b Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8923743.
Location/Qualifiers
FEATURES
source
1..772
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:55585"
/db_xref="taxon:9606"
/clone="MGC:21081 IMAGE:4151953"
/tissue_type="Brain, glioblastoma with EGFR amplification"
/clone_lib="NCL CGAP_Brn64"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
88..303
/codon_start=1
/product="NICE-5 protein"
/protein_id="AAH15316.1"
/db_xref="GI:15929787"
/translation="MELLFKQGHSSAYSIESVIMQISATLVKARVQFGANKSQYSL
TRAQSQYKSLVQIIEKNGWYTPPKEDG"
BASE COUNT 220 a 187 c 157 g 208 t
ORIGIN
Query Match 34.3%; Score 772; DB 9; Length 772;
Best Local Similarity 100.0%; Pred. No. 3.9e-86;
Matches 772; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1483 GATAACTTCCCTTTGACCCACCATTTGTCCAGGTTGTGTCTCCAGTCTCTCGAGGG 1542
DB 1 GATAACTTCCCTTTGACCCACCATTTGTCCAGGTTGTGTCTCCAGTCTCTCGAGGG 60
QY 1543 TATGTTCTGGCGGAGGGCCATCTGCATGGAATCTTCTACCAACAGGGCTGGAGCAGT 1602
DB 61 TATGTTCTGGCGGAGGGCCATCTGCATGGAATCTTCTACCAACAGGGCTGGAGCAGT 120
QY 1603 GCCTACTCCATAGATCAGTGCATGATCAGATCAGTGCCTGCTGGTGAAGGGGAAAGCA 1662
DB 121 GCCTACTCCATAGATCAGTGCATGATCAGATCAGTGCCTGCTGGTGAAGGGGAAAGCA 180
QY 1663 CGAGTGCAGTTTGGAGCCAAACAAATCTCAATACAGTCTGACAAGACACAGCAGTCTCTAC 1722
DB 181 CGAGTGCAGTTTGGAGCCAAACAAATCTCAATACAGTCTGACAAGACACAGCAGTCTCTAC 240
QY 1723 AAGTCTCTGTGTCAGATCTCCAGCAAAAAGAGGCTGTGTCACACCCCAAGAGAGCGGC 1782
DB 1723 AAGTCTCTGTGTCAGATCTCCAGCAAAAAGAGGCTGTGTCACACCCCAAGAGAGCGGC
```

241	DB	AAGTCCTTGGTGACATCCACGAAAAAAGCGCTGGTACACACCCCAAAAGAACGGC	300
1783	QY	TAACCCCTGGAGTATCACCCCTTCCTCCCTCCCAAGGCACCACTGGACCAATTAACCTTTGAA	1842
301	DB	TAACCCCTGGAGTATCACCCCTTCCTCCCTCCCAAGGCACCACTGGACCAATTAACCTTTGAA	360
1843	QY	TGCTGTATTTGGGATCTCAGCGTGCCTCTGTGGTTCCCTCCCTCAATTTTCTCGACGTGA	1902
361	DB	TGCTGTATTTGGGATCTCAGCGTGCCTCTGTGGTTCCCTCCCTCAATTTTCTCGACGTGA	420
1903	QY	TAGCTCTGCCTATTGCAAGGCAATGATGGCTATTCTAAACGCTAAGGAAAAAACAAC	1962
421	DB	TAGCTCTGCCTATTGCAAGGCAATGATGGCTATTCTAAACGCTAAGGAAAAAACAAC	480
1963	QY	ACAGAACTGTTTCAAGTACTCAAGACTGACTTTACAGACCAACCAACCACTGCTGGAAC	2022
481	DB	ACAGAACTGTTTCAAGTACTCAAGACTGACTTTACAGACCAACCAACCACTGCTGGAAC	540
2023	QY	CCTTGCTAGCAGGCAATCTTATAAAGAAACTTTTCGAGCCTCCTTATATGCTGGAACT	2082
541	DB	CCTTGCTAGCAGGCAATCTTATAAAGAAACTTTTCGAGCCTCCTTATATGCTGGAACT	600
2083	QY	CAGCTGTGCTCCAGACTAGAGCCTCCTTACCTATGCTATGGATTTTAAATTTATTTCTC	2142
601	DB	CAGCTGTGCTCCAGACTAGAGCCTCCTTACCTATGCTATGGATTTTAAATTTATTTCTC	660
2143	QY	TTATTTTCATGTACACTGCTTTTTTTGGTTTACAGTGTATGATGTGATGTAAGAAAAAT	2202
661	DB	TTATTTTCATGTACACTGCTTTTTTTGGTTTACAGTGTATGATGTGATGTAAGAAAAAT	720
2203	QY	GTATCTTTGGGAAAACAATTACAGTTTGGTTTAAATTTGAAAAAATAAAAAA	2254
721	DB	GTATCTTTGGGAAAACAATTACAGTTTGGTTTAAATTTGAAAAAATAAAAAA	772

RESULT 15

BC051487	BC051487	853 bp	mRNA	linear	ROD 01-MAY-2003
LOCUS	Mus musculus mRNA similar to RIKEN CDNA 2310012M18 gene (cDNA clone IMAGE:1363390).				
DEFINITION					
ACCESSION	BC051487				
VERSION	BC051487.1	GI:30186069			

KEYWORDS
SOURCE
Mus musculus (house mouse)

ORGANISM	MUS	MUSCULUS
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		
61		
62		
63		
64		
65		
66		
67		
68		
69		
70		
71		
72		
73		
74		
75		
76		
77		
78		
79		
80		
81		
82		
83		
84		
85		
86		
87		
88		
89		
90		
91		
92		
93		
94		
95		
96		
97		
98		
99		
100		

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 853)

REFERENCE
AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altshul, S.F., Zeeberg, B., Buetow, K., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavante, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Saha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullay, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.W., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL
OF
CELLULAR PHYSIOLOGY

MEDLINE 22388257

PUBMED
12477932

REFERENCE 2 16500 1 40 0531
12477352

REFERENCE
AUTHORS
Straussberg, P.
2 (pages 1 to 853)

AUTHORS **Strausberg, R.**

TITLE	REMARK
JOURNAL	COMMENT
<p>Direct Submission</p> <p>Submitted (28-APR-2003), National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA</p> <p>NIH-MGC Project URL: http://mgc.nci.nih.gov</p> <p>Contact: MGC help desk</p> <p>Email: cgabs-i@mail.nih.gov</p> <p>Tissue Procurement: Marcello Bento Soares, Ph.D.</p> <p>cDNA Library Preparation: M. Bento Soares, University of Iowa</p> <p>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)</p> <p>DNA Sequencing by: Institute for Systems Biology</p> <p>http://www.systemsbiology.org</p> <p>contact: amadanesystemsbiology.org</p> <p>Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting</p> <p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov</p> <p>Series: IRAK Plate: 113 Row: c Column: 21</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.</p>	

FEATURES
SOURCE

```

1. .853
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image:1363390"
/tissue_type="Thymus gland
/clone_lib="Soares_thymus_
/lab_host="DH10B"
/note="vector: pT73D-Pac"
x1. .369

```

CD5

```

<!--359
/codon_start=1
/product="Similar to RIKEN CDNA 2310012M18 gene"
/protein_id="AAH51487.1"
/db_xref="GI:30277689"
/translation="GTROIKKEGADFIILNFKDNFPDPVPRVVPVLGGYY
LAGGAICMELLFKQWSAYSTESVIMQISATLVKGKARVQFGANKSQYSLTRAQQS
KSLVIOHKENGWYTPPKEDG"
242 a 212 c 180 g 219 t
BASE COUNT
ORIGIN

```

Query Match	30.2%	Score 681.8;	DB 10;	Length 853;
Best Local Similarity	93.3%;	Pred. NO. 5.3e-75;		
Matches 783;	Conservative	O: Mismatches 37;	Indels 19;	Gaps 6;

[illegible]

```
QY 1786 CCTGGAGTATACCCCTTCCCTCCCTCCAGGACCACTGGACCAATTACCTTTGAATGC 1845
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 370 CCTGGAGTCTCACCCTTCCCT-CCCTCCAGGACCACTGGACCAATTACCTTTGAATGC 428
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1846 TGTATTTGGATCTCAGCCTGCCCTCTGTGGTTCCTCCCTCATTTTCCCTGGACGTGATAG 1905
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 429 TGTGTTTGGATCTCAGCCTGCCCTGTG-GTTTCCCTCCCTCATTTTCCCTGGACGTGATAG 487
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1906 CTCTGCCCTATTGGAGGACATGATGGCTATTCTAAAGGCTAAGGAAAAAACA----- 1960
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 CTCTGCCCTATTGGAGGACATGATGGCTATTCTAAAGGCTAAGGAAAAAACAACCCAC 547
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1961 -----ACACAGAACTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACCACTTG 2015
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 AACACACACTGAAGTCTCAAGTACTCAAGACTGACT-----ACAGACCAACCACTCG 602
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2016 CTGGAACCCCTGTAGCAGGCACTTCTTATAAAGAACTTCGAGCCTCCTTATATGCT 2075
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 603 CTGGAACCCCTGTAGCAGGCACTTCTTAT-AAAGAACTCTCAAGCCTCCTTATATGCT 661
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2076 GGAACCTCAGTGTGCTCCAGACTAGAGCCTCCTTACCTATGCTATGGATTTTAATTTA 2135
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 562 GGAACCTCAGTGTGCTCCAGACTAGAGCCTCCTTACCTATGCTATGGGTTTAAATTTA 721
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2136 TTTTCTCTTATTTCATCTACACTGCTTTTGTGTTACAGTGTATGATGGATGTATGA 2195
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 722 TTTTCTCTTATTTCATCTACACTGCTTTTGTGTTACAGTGTATGATGGATGTATG- 780
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2196 AAAAAATGTATCTTTGGAAAAACAATTACAGTTTGTAAATTGAAAAAATAAAAAA 2254
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 AAAAAATGTATCTTTGGAAAAACAATTACAGTTTGTAAATTGAAAAAATAAAAAA 839
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: August 4, 2003, 19:04:22
Job time : 10669 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 12:58:41 ; Search time 740 Seconds
(without alignments)
8222.342 Million cell updates/sec

Title: US-10-005-549-1
Perfect score: 2254
Sequence: 1 ctgcgtctctctacttg.....tttgaataaaaaaa 2254

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2254	100.0	2254	AAD39359	Human RATL1d6 (reg
2	1554.4	69.0	1666	AAH33115	Human colon cancer
3	1329	59.0	1329	AAZ33482	Human prostate can
4	1172.4	52.0	6277	21 AAC76859	Human ORFX ORF2414
5	720	31.9	791	21 AAC75982	Human ORFX ORF1537
6	461.2	20.5	2988	24 ABN59773	Novel human coding
7	432	19.2	432	24 ABN97105	Gene #3603 used to
8	371.4	16.5	390	16 AAT19939	Human gene signatu

c 10	371.4	16.5	870	24	AAD31836	Human pancreatic t
c 11	371.4	16.5	968	21	AAC59573	Human secreted pro
c 12	371.4	16.5	13076	22	AAI06214	Human reproductive
c 13	371.4	16.5	13076	22	AAI07355	Human reproductive
c 14	369.8	16.4	42999	22	ABG65032	Invertebrate forag
c 15	369.8	16.4	1869	22	AAI14006	DNA to infer human
c 16	368.2	16.3	1869	22	AAI14299	Human DNA to infer
c 17	353.4	15.7	42999	25	ABN95733	Gene #2231 used to
c 18	333	14.8	536	22	AAH35089	Human ribosomal RN
c 19	332.6	14.8	362	20	AAV86232	Human colon cancer
c 20	331.4	14.7	1869	22	AAI14005	EST clone S70. Ho
c 21	331.4	14.7	1869	22	AAI14298	DNA to infer mouse
c 22	331.4	14.7	17384	25	ACC44721	Mouse DNA to infer
c 23	331.4	14.7	22118	25	ACC44629	Cosmid pFK161 nucl
c 24	330	14.6	1426	21	AAZ52472	Mouse ribosomal RN
c 25	309.4	13.7	22118	25	AAH11085	Mouse ribosomal RN
c 26	298	13.2	309	22	AAH81601	Human differential
c 27	281	12.5	281	24	ABV89079	Human colon cancer
c 28	280.4	12.4	451	21	AAH30588	Human colon cancer
c 29	276	12.2	1444	23	ABL20409	Drosophila melanog
c 30	276	12.2	2179	23	ABL05399	Drosophila melanog
c 31	269.4	12.0	316	24	ABN93696	Gene #194 used to
c 32	269.4	12.0	316	24	ABL25316	Colon adenocarcino
c 33	260	11.5	980	20	AAZ42109	Human endometrium
c 34	250.4	11.1	278	22	AAH81515	Human differential
c 35	244	10.8	349	22	AAF64725	Novel human polynu
c 36	227.4	10.1	357	24	ABL79430	Human ovarian canc
c 37	226.4	10.0	1510	24	ABO55048	Human ovarian anti
c 38	224.4	10.0	435	20	AAV86060	EST clone C626. H
c 39	220.8	9.8	454	22	AAI24217	Human breast cance
c 40	220.2	9.8	223	24	ABL36952	Human colon tumour
c 41	220	9.8	378	18	AAV04906	Nucleotide sequenc
c 42	220	9.8	378	25	ACC44633	Mouse ribosomal RN
c 43	213	9.4	378	25	ABX11090	Clone no. 161, sub
c 44	210.8	9.4	323	23	ABV62051	Human prostate exp
c 45	194	8.6	411	23	ABX44547	Bovine EST associa

ALIGNMENTS

RESULT 1

AAD39359
ID AAD39359 standard; DNA; 2254 BP.

AC AAD39359;

XX 04-OCT-2002 (first entry)

DE Human RATL1d6 (regulated in activated T-lymphocyte 1d6) gene.

XX Human; ubiquitin conjugating enzyme; UBC; RATL1d6; immune disorder;
KW regulated in activated T-lymphocyte 1d6; neuronal disorder; cancer;
KW tumour; lymphoproliferative; cancer; adenocarcinoma; leukaemia; myeloma;
KW sarcoma; neurodegenerative; inflammation; rheumatoid arthritis; asthma;
KW multiple sclerosis; psoriasis; neuronal; Alzheimer's disease; dementia;
KW depression; epilepsy; acquired immuno deficiency syndrome; allergy;
KW AIDS; anaemia; atopic dermatitis; diabetes mellitus; dermatological;
KW myocardial infarction; renal tubular acidosis; gonadal dysgenesis;
KW dysplasia; cataract; cytostatic; neuroprotective; nontropic; anti-HIV;
KW anticonvulsant; antiinflammatory; Cushing's syndrome; cardiant;
KW ophthalmological; gene; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 517..1785

FT /*tag= a

FT /product= "Human RATL1d6 protein"

XX WO200236741-A2.

PD	10-MAY-2002.
XX	
PF	29-OCT-2001; 2001WO-US46559.
XX	
PR	30-OCT-2000; 2000US-244688P.
XX	
PR	30-JUL-2001; 2001US-308708P.
XX	
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.
XX	
PI	Bowen MA, Wu Y, Yang W, Finger JN;
XX	
DR	WPI; 2002-479758/51.
XX	P-PSDB; AAE24493.
PT	Novel ubiquitin conjugating enzyme polypeptide isolated from activated
PT	human T cell, for screening modulators useful for treating cancer,
PT	immune disorder, lymphoproliferative disorder, neurodegenerative
XX	disorder
XX	
XX	Claim 1; Fig 1; 169pp; English.
XX	
CC	The invention relates to a novel ubiquitin conjugating enzyme (UBC)
CC	homologue, RAL1d6 [regulated in activated T-lymphocytes 1d6] and its
CC	corresponding nucleic acid. The invention also relates to methods for
CC	treating, diagnosing, preventing and screening for disorders related
CC	to the expression of RAL1d6. UBC is useful for screening for candidate
CC	compounds capable of binding to and/or modulating its activity. UBC is
CC	useful for treating an immune or neuronal disorder in a mammal. The
CC	method is useful for treating a cancer or tumour. It is useful for
CC	suppressing the immune response in a subject requiring the suppression.
CC	It is also useful for treating lymphoproliferative disorder, cancer e.g.
CC	adenocarcinoma, leukaemia, myeloma, sarcoma, etc, neurodegenerative
CC	disorder, inflammatory disorders e.g. rheumatoid arthritis, asthma,
CC	multiple sclerosis, psoriasis, etc, neuronal disorders e.g. Alzheimer's
CC	disease, dementia, depression, epilepsy, etc, immune disorder or immune
CC	related disorders such as acquired immuno deficiency syndrome (AIDS),
CC	allergy, anaemia, atopic dermatitis, diabetes mellitus, myocardial
CC	infarction, etc, developmental disorders e.g. Cushing's syndrome, renal
CC	tubular acidosis, gonadal dysgenesis, dysplasia, cataract, etc. The
CC	present sequence is human RAL1d6 gene.
XX	
SQ	Sequence 2254 BP; 547 A; 602 C; 648 G; 457 T; 0 other;
Query Match	100.0%; Score 2254; DB 24; Length 2254;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2254; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1 CTCGCTCTCTTCCTACTTGGATACTGTGGTAATTTCTAGAGCTAAATACATCGCGAGGGGC 60
Db	
QY	1 CTCGCTCTCTCTTCCTACTTGGATACTGTGGTAATTTCTAGAGCTAAATACATCGCGAGGGGC 60
Db	
QY	61 GCTGACCCCCCTTCGCGGGGGGATCGCTGCATTATTCAGATCAAGACCACCGGTCAGC 120
Db	
QY	61 GCTGACCCCCCTTCGCGGGGGGATCGCTGCATTATTCAGATCAAGACCACCGGTCAGC 120
Db	
QY	121 CCCTCTCGGCCCCCGCGCGCGCGCGCGCGCGCGCGCTTTGTGACTCTAGATAACCTC 180
Db	
QY	121 CCCTCTCGGCCCCCGCGCGCGCGCGCGCGCGCGCGCTTTGTGACTCTAGATAACCTC 180
Db	
QY	181 GGGCCCATGCGACGGCCCGCTGCGGGCGAGACCCATTTCGAACGTCCTGCCCTATCAACT 240
Db	
QY	181 GGGCCCATGCGACGGCCCGCTGCGGGCGAGACCCATTTCGAACGTCCTGCCCTATCAACT 240
Db	
QY	241 TTCGATGTGTAGTCGCGCTACCATGTCACCCAGCCGGTGACGGGGATCAGGGTTCGA 300
Db	
QY	241 TTCGATGTGTAGTCGCGCTACCATGTCACCCAGCCGGTGACGGGGATCAGGGTTCGA 300
Db	
QY	301 TTCCGGAGAGGAGCGCTGAGAAGCGGTACCATCCAGAAAGGCAGCAGCGCGCAA 360
Db	
QY	301 TTCCGGAGAGGAGCGCTGAGAAGCGGTACCATCCAGAAAGGCAGCAGCGCGCAA 360
Db	
QY	361 TTACCCATCTCCGGAGGTGCGCGCGGCCCATCTTTCGCGAAGGGGGGATCAGGAAGTCGC 420
Db	

CC	proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, CC bacterial or fungal infection, malaria, autoimmune disorders, asthma, CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance CC coagulation; to inhibit thrombosis; and as a contraceptive.	XX
50	Sequence 6277 BP; 1521 A; 1720 C; 1611 G; 1423 T; 2 other;	
Query Match	52.0%; Score 1172.4; DB 21; Length 6277;	
Best Local Similarity	80.8%; Pred. No. 1.3e-214;	
Matches 1455; Conservative	0; Mismatches 296; Indels 55; Gaps 5;	
QY	447 CGCGCGCGAGCCGCGAGCGCGGAGCGCTCCCGCGCGCGCGCGCGCGGAGCGAGC 506	
DB	2041 CAGACCGCAGGCTTCGCGCGCCCTCGCCATTTTCAGCAGGCGCTCGACGAGCGGAGA 1982	
QY	507 GGAGCGGAGATCGACGAGCGCGCAGCGCGGAGCGGCGAGCAGCAGCGGGGCGGGGAGCA 566	
DB	1981 GCGCGGCCAGGCGCGCGCGGGGCGTGCAGCGCGTGCAGCGCGCTCGGGGCCGGCT 1922	
QY	567 GCTGGGGGCCAGGGGGCGCGCGGGGGCGGGGGCGGCCAGGGGGGGGGCGGGGCC 626	
DB	1921 CCGCTTCCGCGCGCGCTCCCTTCCGCGCGCGCTCCCGCGGAGATGAGGGGAAGATGTC 1862	
QY	627 GGGCGCCTGCTCAGCGGAGAGCTGAAGCTGCTCGAGTCCATCTCCACGCGGCCACGA 686	
DB	1861 CGTGTCAGGCTCAAGCGCGAGCTGAAGTTCCTGGGCTCCTTCGACGAAGCAGCAGCA 1802	
QY	687 GCCTTTCCGATTTGCCAGCGCCCTTCCTGGACGAGCTGAGCTGGCGAGTTCCTGCTGGCTGG 746	
DB	1801 GCGATTCGCGATCGTCAGTTTGAAGCTGGACGAGCTGCAGTGCAGTTCCTGCTGGCGCA 1742	
QY	747 GGCGGAGGGGCGGGGGCGGGGGCGGGCGCGCGAGCGCATCTCCCCCGAGGGGGTCTGGT 806	
DB	1741 GCAGGGCAGCGCCGCATCGCTGCGCGCGCGCATCAGCGC----- 1704	
QY	807 GCCTGGGGATCTCTCGCGATCCAGTCCACATCAGGAGTGCATACCTGCTGTGCCCC 866	
DB	1703 -----TCCACTGCAACATCAGGAATCTATCCATCTTCTTCCACC 1664	
QY	867 CATCTGGTGGTGGTGTGATGACCTTAACCTTGGTGTGTCTTGGAGAGGCTGTGTGA 926	
DB	1663 GATATGGTTTGTGGATTCTGAAGACCCAAATCTGACATCATAGTTCCTGGAAGCTGTAGAAGA 1604	
QY	927 CATAAAGAAGGAATACTCTGTCTATTGGACGATCTCAAGAGGATCATCTCCACCTGTG 986	
DB	1603 TACTAAGAC- ---AACAAATTGCTGCTCAGCAATTAAGTGGTGTGATGTGAACCTCTG 1547	
QY	987 TAAACTCTATAACCTCCCTCAGCATCCAGATCTGGAGATCTGGATCAACCCCTTGCACGC 1046	
DB	1546 CAGTTTATATACTTCTTAAGCACCTGGATGTTGAGATGCTAGATCAACCATACCAC 1487	
QY	1047 AG---AGCAGTGCACAGGAAGACGTGTCTTCAGAA-----GATGAAGATGAGGAGAT 1097	
DB	1486 GGGTCAAGATGGGACACAGAAGAAGTGACTTCAGAAGAAGAGGAAGAAGAAGAGAT 1427	
QY	1098 GCTCTAGGACACAGAAGACTTATGATCACTATGAANTGAAGAGAGAGAGCCACTGAGGG 1157	
DB	1426 GCGTGAAGATATAGAAGACTTATGATCACTATGAGTGAAGGAAGAGAGCCTATTAGTGG 1367	
QY	1158 CAAGAAATCTGAAGATGATGGCATTTGGAAAGAAACTTCGGCCATCTAGAGAAAATTA 1217	
DB	1366 GAAAAGTCAGAGATGAGGAATTTGAAAAGAAATTTGGCAATATTAGAGAAAATTA 1307	
QY	1218 AAGAACCAGAGCAAGATTACTTAATGGTGCAGTGTCTGGCTCGGTGCAGGCCACTGA 1277	
DB	1306 AAAGACTCAAAGGCAAGACCATTTAAATGCTGCAGTGTCTGGCTCGGTGCAGGCCACTGA 1247	
QY	1278 CCGGCTGTGAAGGAGCTCAGGATATATACCGATCACAGATTTCAAAGCGCGGAACCTA 1337	

Db	1246	CCGGCTGATGAAGAGAGCGCTCAGGAGATATATACCAGATCACAGAGTTATATAAACACAGGGGATTTA	1187
Qy	1338	TGCAGTCGAACTCGTGAATGACAGCTGTGTATGATTGGAATGTCAAACTCTCTCAAACTTGA	1397
Db	1186	TTTCTTGGAACTCGTGAATGACAGCTGTGTATGATTGGAATGTAACTCTCTCAAACTTGA	1127
Qy	1398	CCAGGACAGCGCTTTTGCACAACGATCTCCAGATCTCTAAAGAGAAAAGAGAGGCCGACTT	1457
Db	1126	CCAGGACAGCGCTTTTGCACAACGATCTCCAGATCTCTAAAGAGAAAAGAGAGGCCGACTT	1067
Qy	1458	CATTCTACTTAACTTTTCCCTTTTAAAGATAACTTTCCCTTTTGACCCACCAATTTGTCTCAGGGT	1517
Db	1066	CATTCTACTTAACTTTTCCCTTTTAAAGATAACTTTCCCTTTTGACCCACCAATTTGTCTCAGGGT	1007
Qy	1518	TGTGTCTCCAGTCTCTCTCTGAGGGTATGTTCTTGGCGGAGGGGCCATCTGCATGGAACCT	1577
Db	1006	TGTGTCTCCAGTCTCTCTCTGAGGGTATGTTCTTGGCGGAGGGGCCATCTGCATGGAACCT	947
Qy	1578	TCTCACCAACACAGGGCTGGAGCAGTGCTTACTCCATAGAGTCAGTATCATATGACAGATCAG	1537
Db	946	TCTCACCAACACAGGGCTGGAGCAGTGCTTACTCCATAGAGTCAGTATCATATGACAGATCAG	887
Qy	1638	TGCCACACTCGTGAAGGGGAAACACAGAGTGCAGTTTGGAGCCAAACAAATCTCTCAATACAG	1697
Db	886	TGCCACACTCGTGAAGGGGAAACACAGAGTGCAGTTTGGAGCCAAACAAATCTCTCAATACAG	827
Qy	1698	TCTGACAAGAGCACAGCAGTCTTACAAGTCTTGGTGCAGATCCAGG - AAAAAGAGCGCT	1756
Db	826	TCTGACAAGAGCACAGCAGTCTTACAAGTCTTGGTGCAGATCCAGGAGAAAAGAGCGCT	767
Qy	1757	GGTACACACCCCCAAAGAAAGACGGCTAACCCCTGGAGTATCACCCCTTCCCTCCCTCCCCAG	1816
Db	766	GGTACACACCCCCAAAGAAAGACGGCTAACCCCTGGAGTATCACCCCTTCCCTCCCTCCCCAG	707
Qy	1817	GCACCACTGGACCAATTTACCTTTGAATGTGTATTTGGATCTCACCGCTGCTCTGTGGTT	1876
Db	706	GCACCACTGGACCAATTTACCTTTGAATGTGTATTTGGATCTCACCGCTGCTCTGTGGTT	647
Qy	1877	CCCTCCCTCATTTTTCCTGGACGTGATAGCTCTGCCTATTTCAGGACAAATGATGGCTATT	1936
Db	646	CCCTCCCTCATTTTTCCTGGACGTGATAGCTCTGCCTATTTCAGGACAAATGATGGCTATT	587
Qy	1937	CTAAACGCTAAGGAAAAAACAACACAGACTGTTTCAAGTACTCAAGCTGACTTAC	1996
Db	586	CTAAACGCTAAGGAAAAAACAACACAGACTGTTTCAAGTACTCAAGCTGACTTAC	527
Qy	1997	AGACCAACCAACACACTTGTCTGGAACCTTGTAGCAGGCAATCTTATATAAGAAACTTT	2056
Db	526	AGACCAACCAACACACTTGTCTGGAACCTTGTAGCAGGCAATCTTATATAAGAAACTTT	467
Qy	2057	CGAGCCCTCTTATATTGCTGGAAACTCAGCTGTCTCCAGACTAGAGCCCTCCCTTACCTAT	2116
Db	466	CGAGCCCTCTTATATTGCTGGAAACTCAGCTGTCTCCAGACTAGAGCCCTCCCTTACCTAT	407
Qy	2117	GCATAGGATTTTAAATTTTCTCTTATTTGCATGTACACTGCTTTTTTTTGGTTACAGT	2176
Db	406	GCATAGGATTTTAAATTTTCTCTTATTTTCAATGTACACTGCTTTTTTTTGGTTACAGT	347
Qy	2177	GTATGATGGATGTCTATGAAAAAATCTATCTTTGGGAAACAAATTACAGCTTTGTTAATT	2236
Db	346	GTATGATGGATGTCTATGAAAAAATCTATCTTTGGGAAACAAATTACAGCTTTGTTAATT	287
Qy	2237	TGAAAA 2242	
Db	286	TGAAGA 281	
RESULT 5			
AAC75982			
ID	AAC75982 standard; cdna; 791 bp.		
XX			
AC	AAC75982;		
XX			

RESULT 5
AAC75982
ID AAC75982 standard; cDNA; 791 bp.
XX
AC AAC75982;

08-FEB-2001 (first entry)

Human ORF6, polynucleotide sequence

Human; open reading frame; ORF; detection; cytostatic; hepatotropic; vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 9905-0127607.

02-APR-1999; 9905-0127636.

05-APR-1999; 9905-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RÅ, Leach M;

WPI; 2000-602362/57.

P-PSDB; ABA41773.

Novel nucleic acids and peptides derived from open reading frame X,

useful for treating e.g. cancers, proliferative disorders,

neurodegenerative disorders and cardiovascular disease.

Claim 5; Page 2293-2294; 5507pp; English.

AACT4446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORF open reading frames 1 to 3161. The ORF sequences have activities such as: cytostatic; hepatotropic; vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORF-associated disorder. The nucleic acids can be used to express ORF proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 791 BP; 198 A; 197 C; 234 G; 158 T; 4 other;

Query Match

Best Local Similarity 31.9%; Score 720; DB 21; Length 791;

Matches 765; Conservative 0; Mismatches 17; Indels 6; Gaps 3;

QY 630 GCCCTGCTGAGCGGAGAGCTGAAGCTGCTCGAGTCCATCTTCACCGCGCCGACGAGCG 689
DB 3 GCCCTGCTGAGCGGAGAGCTGAAGCTGCTCGAGTCCATCTTCACCGCGCCGACGAGCG 62
QY 690 CTTCCGATTCGCCAGCGCTGCTGGAGAGCTGAGCTGCGAGTTCCTGCTGGCTGGGGC 749
DB 63 CTTCCGATTCGCCAGCGCTGCTGGAGAGCTGAGCTGCGAGTTCCTGCTGGCTGGGGC 122
QY 750 CGGAGGGCCGG 809
DB 123 CGGAGGGCCGG 182
QY 810 TGGGGATCCTGCTCGCATCCACTGCAACATCAGGAGTCAACCTGCTGCTGGCTGGC 869
DB 183 TGGGGATCCTGCTCGCATCCACTGCAACATCAGGAGTCAACCTGCTGCTGGCTGGC 242
QY 870 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 929
DB 243 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
QY 930 AAGAGAGGGAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 989
DB 303 AAGAGAGGGAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362
QY 990 ACTCTATTAACCTCCCTCAGCATCCAGATGCTGGAGATGCTGGATCAACCTTGC 1049
DB 363 ACTCTATTAACCTCCCTCAGCATCCAGATGCTGGAGATGCTGGATCAACCTTGC 422
QY 1050 GCAGTGCACACAGGAG 1109
DB 423 GCAGTGCACACAGGAG 482
QY 1110 AGAGACTTAGATCAGTCAATGAAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1169
DB 483 AGAGACTTAGATCAGTCAATGAAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 542
QY 1170 AGATGATGGCATTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1229
DB 543 AGATGATGGCATTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
QY 1230 GCAAGATTACCTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1289
DB 603 GCAAGATTACCTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 662
QY 1290 GGAGCTCAGG---ATATATACCGATCACAGATTTCAAAGGCGGAACTATGC--AGTC 1344
DB 663 GGAGCTCAGGATATATATACCGATTTCAAAGGCGGAACTATGCAGTTTC 722
QY 1345 GAATCTGT-GAATGACAGTCTGTATGATTGGAATGTCAAAGTCTCAAAGTTGAC 1403
DB 723 GAATCTGTGGAATGACAGTCTGTATGTTGGTGGATGTTCAACTCTCAAAGTTGAC 782
QY 1404 CAGCGCTT 1411
DB 783 CAGCGTTT 790

RESULT 6

ABN59773

ID ABN59773 standard; cDNA; 2988 BP.

AC ABN59773;

XX 28-JUN-2002 (first entry)

DE Novel human coding sequence SEQ ID NO: 184.

XX Human; antianaemic; vulnary; antiinflammatory; immunomodulator;

KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;

KW neuroprotective; antiparkinsonian; protein therapy; EST;

XX expressed sequence tag; gene; ss.

OS Homo sapiens.

DE Invertebrate foraging behaviour associated human DNA sequence #12.
XX Attention deficit hyperactivity disorder; ADHD; hypertensive;
KW invertebrate foraging behaviour; nitric oxide; hypotensive;
KW CGMP-dependent protein kinase; human; neuroleptic; ds;
KW gene therapy.
XX Homo sapiens.
XX WO200259370-A2.
PN 01-AUG-2002.
XX 13-DEC-2001; 2001WO-US48087.
PF 15-DEC-2000; 2000US-0738630.
XX (NEUR-) NEUROSCIENCES RES FOUND INC.
PA Greenspan RJ, Shaw PJ;
PI WPI; 2002-636544/68.
XX Identifying a compound that modulates an attention deficit
PT hyperactivity disorder (ADHD) for treating e.g. ADHD or hypertension,
PT comprises measuring a foraging behaviour an invertebrate -
PS Claim 89; Page 182-195; 246pp; English.
XX The present invention relates to a method of identifying a compound
CC that modulates attention deficit hyperactivity disorder (ADHD) in
CC a mammal. The method comprises administering a test compound to
CC an invertebrate, and measuring a foraging behaviour of the
CC invertebrate, where the compound that modulates the foraging behaviour
CC of the invertebrate is characterised as a compound that modulates
CC ADHD in a mammal. The compound identified by the new method of the
CC invention is useful for diagnosing or treating ADHD, hypertension or
CC other diseases associated with a nitric oxide/CGMP-dependent protein
CC kinase network in a mammal, particularly humans. The method
CC distinguishes a compound that has a specific effect on ADHD,
CC nitric oxide/CGMP-dependent diseases associated with a
CC nitric oxide/CGMP-dependent kinase protein network
CC in a mammal from a compound that has a non-specific effect. The
CC present sequence represents a DNA sequence associated with
CC invertebrate foraging behaviour.
XX
SQ Sequence 42999 BP; 6411 A; 13605 C; 11491 G; 11479 T; 13 other;
Query Match 16.5%; Score 371.4; DB 24; Length 42999;
Best Local Similarity 99.7%; Pred. No. 2.7e-61;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCGCTCTCTCCTACTTGGGATACTGTGGTAATCTAGAGCTAATACATGCGACGGGC 60
|||||
Db 3783 CTCGCTCTCTCCTACTTGGGATACTGTGGTAATCTAGAGCTAATACATGCGACGGGC 3842
QY 61 GCTGACCCCTTCGCGGGGGGATGCGTGCATTTATCATGATCAAGACCAACCCGGTCAAG 120
|||||
Db 3843 GCTGACCCCTTCGCGGGGGGATGCGTGCATTTATCATGATCAAGACCAACCCGGTCAAG 3902
QY 121 CCCTCTCCGCGCCCGCGGGGGGGGGCGGGCGGGCGGGCTTTGGTACTCTAGATAACCTC 180
|||||
Db 3903 CCCTCTCCGCGCCCGCGGGGGGGGGCGGGCGGGCGGGCTTTGGTACTCTAGATAACCTC 3962
QY 181 GGGCGGATCGCAGCCCGCCCTGGGGGGGAGCGACCCATTTCGAACCTCTGCGCTATCAACT 240
|||||
Db 3963 GGGCGGATCGCAGCCCGCCCTGGGGGGGAGCGACCCATTTCGAACCTCTGCGCTATCAACT 4022
QY 241 TTCGATGTAGTGCCTGCGCTTACCATGTTGACACCGGTGACGGGGGAATCAGGGTTTCA 300
|||||
Db 4023 TTCGATGTAGTGCCTGCGCTTACCATGTTGACACCGGTGACGGGGGAATCAGGGTTTCA 4082
QY 301 TTCGCGAGGGAGCGCTGAGAAACGGCTACCAATCCCAAGGAGGACGACGCGCGGCAAA 360

Db 4083 TTCGGGAGAGGGAGCCCTGAGAAACGGCTACCAATCCCAAGGAGGACGACGCGGCAAA 4142
|||||
QY 361 TTACCCACTCCCG 373
|||||
Db 4143 TTACCCACTCCCG 4155
|||||
RESULT 14
AAD14006
ID AAD14006 standard; DNA; 1869 BP.
XX
AC AAD14006;
XX
XX 06-NOV-2001 (first entry)
XX DNA to infer human 18S ribosomal RNA.
XX
XX Translational regulatory element; Internal ribosome entry site; IRES;
KW cistron; protein stability; cell viability; protein yield; toxin;
KW therapeutic agent; human; 18S ribosome; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT modified_base 27 /*tag= a
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
FT modified_base 99 /*tag= b
FT /*mod_base= OTHER
FT /*note= "2'-O-methyladenosine"
FT modified_base 116 /*tag= c
FT /*mod_base= um
FT modified_base 121 /*tag= d
FT /*mod_base= um
FT modified_base 159 /*tag= e
FT /*mod_base= OTHER
FT modified_base 166 /*note= "2'-O-methyladenosine"
FT /*tag= f
FT /*mod_base= OTHER
FT modified_base 172 /*tag= g
FT /*mod_base= um
FT modified_base 174 /*tag= h
FT modified_base 428 /*tag= i
FT /*mod_base= um
FT modified_base 436 /*tag= j
FT /*mod_base= gm
FT modified_base 462 /*tag= k
FT /*mod_base= cm
FT modified_base 468 /*tag= l
FT /*mod_base= OTHER
FT modified_base 484 /*note= "2'-O-methyladenosine"
FT /*tag= m
FT modified_base 509 /*note= "2'-O-methyladenosine"
FT /*tag= n
FT /*mod_base= gm

FT	modified_base	512	
FT			/*tag= o
FT			/mod_base= OTHER
FT	modified_base	517	/note= "2'-O-methyladenosine"
FT			
FT			/*tag= p
FT	modified_base	576	/mod_base= cm
FT			
FT			/*tag= q
FT			/mod_base= OTHER
FT	modified_base	590	/note= "2'-O-methyladenosine"
FT			
FT			/*tag= r
FT			/mod_base= OTHER
FT	modified_base	601	/note= "2'-O-methyladenosine"
FT			
FT			/*tag= s
FT	modified_base	627	/mod_base= gm
FT			
FT			/*tag= t
FT	modified_base	644	/mod_base= um
FT			
FT			/*tag= u
FT	modified_base	668	/mod_base= gm
FT			
FT			/*tag= v
FT			/mod_base= OTHER
FT	modified_base	683	/note= "2'-O-methyladenosine"
FT			
FT			/*tag= w
FT	modified_base	797	/mod_base= gm
FT			
FT			/*tag= x
FT	modified_base	799	/mod_base= cm
FT			
FT			/*tag= y
FT	modified_base	867	/mod_base= um
FT			
FT			/*tag= z
FT	modified_base	1031	/mod_base= gm
FT			
FT			/*tag= aa
FT			/mod_base= OTHER
FT	modified_base	1248	/note= "2'-O-methyladenosine"
FT			
FT			/*tag= ab
FT			/mod_base= OTHER
FT			/note= "3-(3-amino-3-carboxypropyl)-1-methylpseudouridine"
FT	modified_base	1288	
FT			/*tag= ac
FT			/mod_base= um
FT	modified_base	1326	
FT			
FT			/*tag= ad
FT	modified_base	1328	/mod_base= um
FT			
FT			/*tag= ae
FT	modified_base	1383	/mod_base= gm
FT			
FT			/*tag= af
FT			/mod_base= OTHER
FT	modified_base	1391	/note= "2'-O-methyladenosine"
FT			
FT			/*tag= ag
FT	modified_base	1442	/mod_base= cm
FT			
FT			/*tag= ah
FT	modified_base	1447	/mod_base= um
FT			
FT			/*tag= ai
FT	modified_base	1490	/mod_base= gm
FT			
FT			/*tag= aj


```
OY 61 GCTGACCCCTTCGCGGGGGGATGCGTGCATTATCAGATCAAGACCAACCGGTGACG 120
DB 187 GCTGACCCCTTCGCGGGGGGATGCGTGCATTATCAGATCAAAACCAACCGGTGACG 246
OY 121 CCCTCTCGCGCCCGCGGGCGCGCGCGCGCTTTGGTGACTCTAGATAACCTC 180
DB 247 CCCTCTCGCGCCCGCGGGCGCGCGCGCGCTTTGGTGACTCTAGATAACCTC 306
OY 181 GGGCGGATCGACACCCCGCTGCGCGCGGACGCCATTTCGAAGCTCTGCCCTATCAACT 240
DB 307 GGGCGGATCGACACCCCGCTGCGCGGCGGAGGCCATTTCGAAGCTCTGCCCTATCAACT 366
OY 241 TTCGATGTAAGTCCCGCTGCTACCATGGTGACACACGGGTGACGGGAATCAGGGTTCGA 300
DB 367 TTCGATGTAAGTCCCGCTGCTACCATGGTGACACACGGGTGACGGGAATCAGGGTTCGA 426
OY 301 TTCGGGAGAGGACCTTGAGAACGGCTACACATCCAAAGGAGGAGCGCGCGGAAA 360
DB 427 TTCGGGAGAGGACCTTGAGAACGGCTACACATCCAAAGGAGGAGCGCGCGGAAA 486
OY 361 TTACCACTCCCG 373
DB 487 TTACCACTCCCG 499

RESULT 15
AADI4299
ID AADI4299 standard; DNA; 1869 BP.
XX AADI4299;
XX
XX 06-NOV-2001 (first entry)
XX Human DNA to infer 18S ribosomal RNA.
DE
XX
KW Transcriptional regulatory element; translational regulatory element;
KW gene expression; diagnostic application; therapeutic application;
KW human; ribosomal RNA; 18S rRNA; internal ribosome entry site; IRES; ds.
XX
XX Homo sapiens.
XX
FH key Location/Qualifiers
FT modified_base 27
FT /*tag= a
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 99
FT /*tag= b
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 116
FT /*tag= c
FT /mod_base= um
FT modified_base 121
FT /*tag= d
FT /mod_base= um
FT modified_base 159
FT /*tag= e
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 166
FT /*tag= f
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 172
FT /*tag= g
FT /mod_base= um
FT modified_base 174
FT /*tag= h
FT /mod_base= cm
FT modified_base 428
FT /*tag= i
FT /mod_base= um
```

```
FT modified_base 436
FT /*tag= j
FT /mod_base= gm
FT modified_base 462
FT /*tag= k
FT /mod_base= cm
FT modified_base 468
FT /*tag= l
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 484
FT /*tag= m
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 509
FT /*tag= n
FT /mod_base= gm
FT modified_base 512
FT /*tag= o
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 517
FT /*tag= p
FT /mod_base= cm
FT modified_base 576
FT /*tag= q
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 590
FT /*tag= r
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 601
FT /*tag= s
FT /mod_base= gm
FT modified_base 627
FT /*tag= t
FT /mod_base= um
FT modified_base 644
FT /*tag= u
FT /mod_base= gm
FT modified_base 668
FT /*tag= v
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 683
FT /*tag= w
FT /mod_base= gm
FT modified_base 797
FT /*tag= x
FT /mod_base= cm
FT modified_base 799
FT /*tag= y
FT /mod_base= um
FT modified_base 867
FT /*tag= z
FT /mod_base= gm
FT modified_base 1031
FT /*tag= aa
FT /mod_base= OTHER
FT /note= "2'-O-methyladenosine"
FT modified_base 1248
FT /*tag= ab
FT /mod_base= OTHER
FT /note= "3-(3-amino-3-carboxypropyl)-
1-methylpseudouridine"
FT modified_base 1288
FT /*tag= ac
FT /mod_base= um
FT modified_base 1326
FT /*tag= ad
FT /mod_base= um
FT modified_base 1328
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 13:47:14 ; Search time 52 Seconds

(without alignments)
963.781 Million cell updates/sec

Title: US-10-005-549-2

Perfect score: 2229

Sequence: 1 MQQPQGOQQPGPQQQLGG.....KSLVQIHEKNGWYTPPKEDG 422

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:**

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1863	83.6	374	15	US-10-106-698-4458
2	140.5	6.3	46	9	US-09-864-761-43956
3	137	6.1	165	9	US-09-842-528-4
4	135.5	6.1	295	15	US-10-060-019-22
5	134.5	6.0	223	15	US-10-128-714-8332
6	133	6.0	238	15	US-10-128-714-3352
7	122.5	5.5	1048	15	US-10-174-363-56
8	121.5	5.5	144	9	US-09-842-528-5
9	121.5	5.5	1046	15	US-10-156-761-10088
10	120.5	5.4	1004	10	US-09-738-626-5676
11	119.5	5.4	256	9	US-09-925-301-1083
12	118	5.3	527	10	US-09-738-973-216
13	118	5.3	527	10	US-09-854-133-216
14	118	5.3	527	15	US-10-144-649A-216
15	116.5	5.2	1786	10	US-09-742-096-3

16	116	5.2	1647	11	US-09-824-574-4	Sequence 4, Appli
17	115.5	5.2	265	11	US-09-975-719-31	Sequence 31, Appl
18	114	5.1	519	15	US-10-113-794A-2	Sequence 2, Appli
19	113.5	5.1	176	15	US-10-043-487-401	Sequence 401, App
20	112.5	5.0	161	9	US-09-925-297-801	Sequence 801, App
21	111.5	5.0	147	15	US-10-153-668-292	Sequence 292, App
22	111.5	5.0	147	15	US-10-153-668-406	Sequence 406, App
23	111.5	5.0	525	15	US-10-168-425-7	Sequence 7, Appli
24	109.5	4.9	147	14	US-10-108-605-331	Sequence 331, App
25	109	4.9	1084	15	US-10-217-700-9	Sequence 9, Appli
26	108	4.8	592	14	US-10-047-593-2	Sequence 2, Appli
27	108	4.8	592	14	US-10-047-593-4	Sequence 4, Appli
28	107.5	4.8	147	15	US-10-153-668-294	Sequence 294, App
29	107.5	4.8	147	15	US-10-153-668-408	Sequence 408, App
30	107.5	4.8	148	9	US-09-826-312-4	Sequence 4, Appli
31	107.5	4.8	148	15	US-10-108-767-4	Sequence 4, Appli
32	107.5	4.8	148	15	US-10-152-156-4	Sequence 4, Appli
33	107	4.8	148	15	US-10-043-487-412	Sequence 412, App
34	106.5	4.8	143	15	US-10-043-487-399	Sequence 399, App
35	106.5	4.8	641	11	US-09-959-987-2	Sequence 2, Appli
36	106.5	4.8	647	9	US-09-925-299-1002	Sequence 1002, Ap
37	106.5	4.8	647	11	US-09-925-299-1002	Sequence 1002, Ap
38	105.5	4.7	143	15	US-10-043-487-392	Sequence 392, App
39	105.5	4.7	462	11	US-09-919-039-324	Sequence 324, App
40	105.5	4.7	1070	15	US-10-128-714-8314	Sequence 8314, Ap
41	105	4.7	510	9	US-09-866-362-56	Sequence 56, Appl
42	104.5	4.7	151	9	US-09-842-528-2	Sequence 2, Appli
43	104.5	4.7	151	15	US-10-153-668-150	Sequence 150, App
44	103.5	4.6	758	10	US-09-903-248-2	Sequence 2, Appli
45	103.5	4.6	758	10	US-09-859-604-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-106-698-4458
; Sequence 4458, Application US/10106698
; Publication No. US20030109690A1

; GENERAL INFORMATION:
; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept

; FILE REFERENCE: PRA005P1

; CURRENT APPLICATION NUMBER: US/10/106,698

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: PCT/US00/26524

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING DATE: 1999-09-29

; PRIOR APPLICATION NUMBER: US 60/163,280

; PRIOR FILING DATE: 1999-11-03

; NUMBER OF SEQ ID NOS: 8564

; SOFTWARE: PatentIn Ver. 3.0

; SEQ ID NO 4458

; LENGTH: 374

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: MISC_FEATURE

; LOCATION: (25)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: MISC_FEATURE

; LOCATION: (38)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-106-698-4458

Query Match 83.6%; Score 1863; DB 15; Length 374;

Best Local Similarity 98.4%; Pred. No. 3.1e-132;

Matches 359; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 59 FRTASACLDSECEFLIAGAGAGAGAAFGPHLPFRGSGVDPVRHICNTESYPAPVPI 118

|||||

Db 11 FRTASACLDSECEFLIAGAGAGAGAXPGTASPTGSGVDPVRHICNTESYPAPVPI 70

QY 96 -----SVPGDPVR-----IHCNITESPAY--PPIWSVEDDPNLAAY-LEP--- 134
Db 400 PGGGGTQAGFGRPGARRGRKSKRQRRQEAQAQSVGGVMLPRNGQSVLRSGAS 459
QY 135 LVDIKKGWTLLOHLKRIISDLCKLYNLPQHPDVEMLDQPLPAEQCTQEDVSSDEDEEM 194
Db 460 LTDFAEIKGANPASLVGMNMLGEMVTATQSVSDETLKLADENMFVLEIVSPEEDREL 519
QY 195 PEDTEDLDHYEMKEEPAE-----GKKSDEDDGIGKENLAILEIKKN 236
Db 520 LESFD-----IFEGDEGEELVAPPPVTVMGVHDGKTRLLDTIRKTNVAGEA----- 571
QY 237 QRODYLGAVS---GSVQATDRMLKELRDI-----YRSQFSGNGNVAVELVN 280
Db 572 -----GGITQHIGAYQVTFEVNDEERKIFIDTPGHEATAMRAKAKSTDAI----- 620
QY 281 DSLYDWNVKKLVKVDQSDALHNDLQILKEKEGADF-ILLNFSKDNFPDPFVRYVSVPL 339
Db 621 -----LVVAANDGVMPOTIEALNHAKAADVPVIVAVNKKIDVEGADPTKVR--GOLT 669
QY 340 SGGVY---LGGGAICMELLTKQGSNAYSIESVINQISATL 377
Db 670 EFLVAEYGGDTMEFVDSIAKQGLNIESILLEAVLTADASL 710

RESULT 10

US-09-738-626-5676

; Sequence 5676, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 5676
; LENGTH: 1004
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5676

Query Match 5.4%; Score 120.5; DB 10; Length 1004;
Best Local Similarity 21.7%; Pred. No. 1.6;
Matches 97; Conservative 46; Mismatches 112; Indels 191; Gaps 24;
QY 4 PQPGQQQPGGQQQLGGGGA-----AP---GAGGGPGG 33
Db 281 PRPGGSRPGGSAQAQAQAPSQERQGGRRPSPAMPPPTPQMPAKAPGKGRGQAGG 340
QY 34 GPGGPGCLRRLLKLLLESIFHRGHERFIASACLDLSCEFFLAGGAGAGAAAPGPHLPP 93
Db 341 GAGGG-----FNRG-----GGTGGGAGRGR--- 361
QY 94 RGSVPG-----DPVRIHCNITESPAYPPPIWSVEDDPNLAAYLERLVDIKK 140
Db 362 RGGTAGAGRPGGAPRRGRKSKRQRRQEAQAQSVGGVMLPRNGQSVLRSGAS 408

QY 141 GNTL-----LLOHLKRIISDLCKLYNLPQ--HPDVEMLDQPLPAEQCTQED--- 184
Db 409 GATIRLAGASLADFADKIGADAAALVOALFNLGEMVTATASVSDETL---QLLGEEMNY 465
QY 185 ---VSSDEDEEMPEDTEDLDHYEMKEEPAEKGKSEDDGIGKENLA----- 228
Db 466 KVQVVSPEDEDEL-----LESFDL-----QFGEDEG-GEADLAKRPPVTVMGHV 510
QY 229 -----ILEIKK-NORODYLNGAVSG-----SVOATDRML-----KELRDIYRS 266
Db 511 DHGKTRLLDTIRKANVGSDEAGGTQIGAGYQVKNVVEDTERTITFTDTPGHEAFTAMRA 570
QY 267 QSFKGGNVAVELV--NDSLYDWNVKKL-----KVDQSDALHNDLQ----- 304
Db 571 RGAKSTDIADVLAADGVMPQTVTEALNHAKAADVPVIVAVNKKIDKPEASPEKIRGOLTE 630
QY 305 --ILKEKEGADFILLNFSKDNFPDF 328
Db 631 YGLIPEYGGDTIFVDSIAKQGLNID 656

RESULT 11

US-09-925-301-1083

; Sequence 1083, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 1083
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (56)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1083

Query Match 5.4%; Score 119.5; DB 9; Length 256;
Best Local Similarity 22.9%; Pred. No. 0.29;
Matches 49; Conservative 32; Mismatches 96; Indels 37; Gaps 8;
QY 187 SEDEDEEMPEDTEDLDHYEMKEEPAEKGKSEDDGIGKENLAILEIKKNQRODYLGAV 246
Db 65 SDDSRASTSSSSSSSSSQTEKETNTPKKKESVSKSKLLSTSAK----- 113
QY 247 SGVQATDRMLKELRDIYRSQSFKGGNVAVELNDSLYDNVKKLVKVDQSDALHNDLQIL 306
Db 114 -----RIQELADITLDP-----PNCAGPKGDNIEWRSTIL--GPFGSVY----- 154
QY 307 KEKGADEPILLNFSKDNFPDPFVRYVSVPLSGGVGLGGGAICMELLTKQGSNAYS 366
Db 155 ---EGGVF-FIDITFTPEYFPKPKVTRFRIYHCN-INSQGVICLDIL-KDNWSPALTI 208
QY 367 ESVMQISATLVK-GKARVQFGANKSOYSLTRAQ 399
Db 209 SKVLLSCSLTDCNPADPLVGSITATQYMTNRAE 242

RESULT 12

US-09-738-973-216

; Sequence 216, Application US/09738973
; Patent No. US20020110563A1


```
QY 40 CLRRELKLEIFRHERFRIA-----SACLDLSCFLLAGAGAGAPGHP-- 92
Db : : | | : | : | : | : | : | : | : | : | : | : | : |
220 PPEAOPRLEPAPRA--RPEVPEGEPCADSRAGGDTALSGDDPKPKPKKGPMPRL 277
QY 93 --PRGSVPDPVRIHCNITESPAYVPIWSVESDDPNLAAVLERLVLDIKKGNLTLLQHLK 150
Db : : | : | : | : | : | : | : | : | : | : | : | : | : |
278 FLDLGPQGNSEQIKARLSRLSLAPLTLTPFPFGP----- 314
QY 151 RIISDLCKLYNLPOHPDVEMLD-----OPLPAEQCTQEDVSSEDEDEMPEDT 198
Db : : | : | : | : | : | : | : | : | : | : | : | : | : |
315 -----PRPPPEGADAGAAGGAGAGAPGPAE-----EDGEDEDEE- 353
QY 199 EDLDHYEMKEEPAEGKSEDDGIGKENLAILEIKKKNORODY---LNGAVSGSVQATDR 255
Db : : | : | : | : | : | : | : | : | : | : | : | : | : |
354 ED-----EAAAFGAAGPGRGARAAPVPPVVSSADADAARPLRGLLKSPRGADEP 406
QY 256 LMKELDRYRSQSFKGGNYAVELVNDSLYDMNVKLLKVDQSALHNDLIQLKEG----- 311
Db : : | : | : | : | : | : | : | : | : | : | : | : | : |
407 EDELERKRMVSPHG-----DVTYLFDOETPT-NELSVQAPPEGDTDP 450
QY 312 -----ADFILLN-----FSKDNFPDPP-----FVRV-VSPVL 339
Db : : | : | : | : | : | : | : | : | : | : | : | : | : |
451 STPPAPTPPHPATPGDGFPSNDSGFGSFEWAEDFLLPPPGPLCFRFSVSPAL 507
```

RESULT 15

```
US-09-742-096-3
; Sequence 3, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 2007/3050DIV
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: P. falciparum
US-09-742-096-3
```

```
Query Match 5.2%; Score 116.5; DB 10; Length 1786;
Best Local Similarity 25.7%; Pred. No. 6.8;
Matches 58; Conservative 36; Mismatches 79; Indels 53; Gaps 10;

QY 167 DVEMLDQPLPAEOCTQEDVSSEDEDEMPEDTDLHYEMKEEPAEGKSEDDGCI-GKE 225
Db : | | : : | : | : | : | : | : | : | : | : | : | : | : | : |
1532 EYEMKDEDI--EEDVEEDIEEDIEEDKVEDDEDID-----EDIGEDKDEVIDLIQKE 1583

QY 226 NLAILEKIKKNQODLYNGAVSGVQATDRMLKELDRYRSQSFKGGNYAVELVNDSLYD 285
Db : | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
1584 KRIEKVKAKKKLEKKEVEEYSGLKKHVDYMKYVQKIDKVD-KEVSKALESKNDV--- 1639

QY 286 WNVKLLKVDQSALHNDLIQLKEGADFLLNFSKDNF-----PDPFPFVRYVSPVLS 340
Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
1640 --TNVLKQKQDF-----FSKVNKFNKVKYKVAAPFISAVAAPAS 1676

QY 341 GGYVLGGGAICMELLTKOGWSAYSIES---VIMQISATLYKGRAR 383
Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
1677 --YVVG-----FFTSLSFSSCVTIASSTYLLSKVKDTINKNER 1713
```

Search completed: July 29, 2003, 13:56:17
Job time : 54 secs

RESULT 3
US-08-093-144-6
; Sequence 6, Application US/08093144
; Patent No. 5434048
; GENERAL INFORMATION:
; APPLICANT: SIMON, LUC
; APPLICANT: LAONDE, MAURICE
; TITLE OF INVENTION: DNA PROBES FOR THE DETECTION OF
; NUMBER OF INVENTION: ARBUSCULAR ENDOMYCORRHIZAL FUNGI-
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

```

RESULT 4
US-08-998-416-952/c
Sequence 952, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgan
APPLICANT: Knechtle, Philipp
APPLICANT: Rebschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF THE HUMAN
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Cor
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park

```

```

; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 952:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1588UP
; US-08-998-416-952
;
; Query Match
; Best Local Similarity 6.1%; Score 137.2; DB 3; Length 723;
; Matches 169; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
;
; QY 150 GCCGCGGCTTTGGTGAATCTAGATAACCTCGGGCGGATCGGCACGCCGCCCGTGGCGGG 209
; DB 454 GTCTTCGGACTCTTTGATGATTCAATAAATTTTCGAATCGATCGCTTGTCTGGCG 395
;
; QY 210 AGCACCATTGCAACCTCTGCCCTATCAATTCGATGTAGTCCGCGCTACCATGG 269
; DB 394 ATGGTTCATCAATTTCTGCCCTATCAATTCGATGTAGTGGCTTACCATGG 335
;
; QY 270 TGACACCGGTGACGGGGAATCAGGGTTCGATTCGGAGAGGGAGCCTGAGAAACGGCTA 329
; DB 334 TTTCACCGGTGTAACGGGGAATAAGGTTTCGATTCGGAGAGGGAGCCTGAGAAACGGCTA 275
;
; QY 330 CCACATCCAAGGAGCAGCGCGGCGGCAAAATTAACCACTCC 371
; DB 274 CCACATCCAAGGAGCAGCGCGGCGGCAAAATTAACCACTCC 233
;
; RESULT 5
; US-09-350-710B-1/c
; Sequence 1, Application US/09350710B
; Patent No. 6489158
; GENERAL INFORMATION:
; APPLICANT: Hendrick, Carol A.
; APPLICANT: Platt, Nancy J.
; APPLICANT: Ruser, Barbara G.
; APPLICANT: Hoganson, Dean A.
; TITLE OF INVENTION: No. 6489158-Lactate Assimilating Yeast for Improving Aerobic
; FILE REFERENCE: Stability of Silage
; CURRENT APPLICATION NUMBER: US/09/350,710B
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1776
; TYPE: DNA
;
; ORGANISM: Yeast integration vector
; US-09-350-710B-1
;
; Query Match
; Best Local Similarity 6.1%; Score 137.2; DB 4; Length 1776;
; Matches 169; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
;
; QY 150 GCCGCGGCTTTGGTGAATCTAGATAACCTCGGGCGGATCGGCACGCCGCCCGTGGCGGG 209
; DB 1550 GTCTTCGGACTCTTTGATGATTCAATAAATTTTCGAATCGATCGCTTGTCTGGCG 1491
;
; QY 210 AGCACCATTGCAACCTCTGCCCTATCAATTCGATGTAGTCCGCGCTACCATGG 269
; DB 1490 ATGGTTCATCAATTTCTGCCCTATCAATTCGATGTAGTGGCCCTACCATGG 1431
;
; QY 270 TGACACCGGTGACGGGGAATCAGGGTTCGATTCGGAGAGGGAGCCTGAGAAACGGCTA 329
; DB 1430 TTTCACCGGTGTAACGGGGAATAAGGTTTCGATTCGGAGAGGGAGCCTGAGAAACGGCTA 1371
;
; QY 330 CCACATCCAAGGAGCAGCGCGGCGGCAAAATTAACCACTCC 371
; DB 1370 CCACATCCAAGGAGCAGCGCGGCGGCAAAATTAACCACTCC 1329
;
; RESULT 6
; US-08-998-416-178/c
; Sequence 178, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtel, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 178:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 703 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:

```


Db 120 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 61
Qy 345 GCAGCAGCGCGCAAAATACCACATCCCG 373
Db 60 GCAGCAGCGCGCAAAATACCACATCCCG 32

RESULT 9

US-08-592-541-19/c
; Sequence 19, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592.541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-592-541-19

Query Match 6.0%; Score 135.4; DB 2; Length 395;
Best Local Similarity 98.7%; Pred. No. 7.3e-17;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 226 TCTGCCCTATCAACTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 284
Db 180 TCTGCCCTATTAACCTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 121
Qy 285 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 344
Db 120 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 61
Qy 345 GCAGCAGCGCGCAAAATACCACATCCCG 373
Db 60 GCAGCAGCGCGCAAAATACCACATCCCG 32

RESULT 10

US-09-124-698-19/c
; Sequence 19, Application US/09124698
; Patent No. 6117978
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M

; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124.698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592.541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-124-698-19
Query Match 6.0%; Score 135.4; DB 3; Length 395;
Best Local Similarity 98.7%; Pred. No. 7.3e-17;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 226 TCTGCCCTATCAACTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 284
Db 180 TCTGCCCTATTAACCTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 121
Qy 285 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 344
Db 120 GGGATCAGGTTTCGATTCGGAGGAGGAGGCTGAGAAACGGCTACCATCCCAAGGAAG 61
Qy 345 GCAGCAGCGCGCAAAATACCACATCCCG 373
Db 60 GCAGCAGCGCGCAAAATACCACATCCCG 32

RESULT 11

US-09-127-480-19/c
; Sequence 19, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/127,480
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-127-480-19

Query Match 6.0%; Score 135.4; DB 3; Length 395;
Best Local Similarity 98.7%; Pred. No. 7.3e-17;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 226 TCTGCCCTATCAACTTCGATGTCGGCGTGCT-ACCATGGTGACACGGGTGACG 284
|||||
DB 180 TCTGCCCTATTAACCTTCGATGTCGGCGTGCTAACCATGGTGACACGGGTGACG 121
QY 285 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCTGAGAAACGGCTACCATCCCAAGGAAG 344
|||||
DB 120 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCTGAGAAACGGCTACCATCCCAAGGAAG 61
QY 345 GCAGCAGCGCGCGCAAAATTACCCACTCCCG 373
|||||
DB 60 GCAGCAGCGCGCGCAAAATTACCCACTCCCG 32

RESULT 12
US-08-496-841C-19/C
Sequence 19, Application US/08496841C
Patent No. 6210919
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
ROMMENS, JOHANNA M
FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby, PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,841C
FILING DATE: 28-Jun-1995
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul F. Fehlner, Ph.D.
REGISTRATION NUMBER: 35,135
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-496-841C-19
Query Match 6.0%; Score 135.4; DB 3; Length 395;
Best Local Similarity 98.7%; Pred. No. 7.3e-17;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 226 TCTGCCCTATCAACTTCGATGTCGGCGTGCT-ACCATGGTGACACGGGTGACG 284
|||||
DB 180 TCTGCCCTATTAACCTTCGATGTCGGCGTGCTAACCATGGTGACACGGGTGACG 121
QY 285 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCTGAGAAACGGCTACCATCCCAAGGAAG 344
|||||
DB 120 GGAATCAGGGTTCGATTCGGAGAGGAGGAGCTGAGAAACGGCTACCATCCCAAGGAAG 61
QY 345 GCAGCAGCGCGCGCAAAATTACCCACTCCCG 373
|||||
DB 60 GCAGCAGCGCGCGCAAAATTACCCACTCCCG 32

RESULT 13
US-09-124-523-19/C
Sequence 19, Application US/09124523
Patent No. 6395960
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
ROMMENS, JOHANNA M
FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-124-523-19

Query Match 6.0%; Score 135.4; DB 4; Length 395;
Best Local Similarity 98.7%; Pred. No. 7.3e-17;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 226 TCTGCCCTATCAACTTCGATGTAGTCCGCGCT-ACCATGTGACCGGGTGACG 284
|||||
DB 180 TCTGCCCTATTAACTTCGATGTAGTCCGCGCTAACCATGTGACCGGGTGACG 121
|||||
QY 285 GGAATCAGGGTTCGATTCGGAGAGGAGCGCTGAGAAACGGCTACCATCCAGGAAG 344
|||||
DB 120 GGAATCAGGGTTCGATTCGGAGAGGAGCGCTGAGAAACGGCTACCATCCAGGAAG 61
|||||
QY 345 GCAGCAGCGCGCAAAATTACCCACTCCCG 373
|||||
DB 60 GCAGCAGCGCGCAAAATTACCCACTCCCG 32
|||||

RESULT 14

US-09-636-796A-19/c
; Sequence 19, Application US/09636796A
; Patent No. 6485911

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H
; ROMMENS, JOHANNA M
; FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/636,796A

FILING DATE: 11-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/592,541

FILING DATE: <Unknown>

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-636-796A-19

Query Match 6.0%; Score 135.4; DB 4; Length 395;
Best Local Similarity 98.7%; Pred. No. 7.3e-17;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 226 TCTGCCCTATCAACTTCGATGTAGTCCGCGCT-ACCATGTGACCGGGTGACG 284
|||||
DB 180 TCTGCCCTATTAACTTCGATGTAGTCCGCGCTAACCATGTGACCGGGTGACG 121
|||||
QY 285 GGAATCAGGGTTCGATTCGGAGAGGAGCGCTGAGAAACGGCTACCATCCAGGAAG 344
|||||
DB 120 GGAATCAGGGTTCGATTCGGAGAGGAGCGCTGAGAAACGGCTACCATCCAGGAAG 61
|||||

QY 345 GCAGCAGCGCGCAAAATTACCCACTCCCG 373
|||||
DB 60 GCAGCAGCGCGCAAAATTACCCACTCCCG 32
|||||

RESULT 15

US-08-431-048F-19/c

; Sequence 19, Application US/08431048F

; Patent No. 6531586

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H
; ROMMENS, JOHANNA M
; FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 155

CORRESPONDENCE ADDRESS:

ADDRESSEE: DARBY & DARBY P.C.

STREET: 805 THIRD AVENUE

CITY: NEW YORK

STATE: N.Y.

COUNTRY: U.S.A.

ZIP: 10022-7513

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/431,048F

FILING DATE: 28-Apr-1995

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: FEHLNER, PAUL F.

REGISTRATION NUMBER: 35135

REFERENCE/DOCKET NUMBER: 1034/0F808

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-527-6237

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-08-431-048F-19

Query Match 6.0%; Score 135.4; DB 4; Length 395;
Best Local Similarity 98.7%; Pred. No. 7.3e-17;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 226 TCTGCCCTATCAACTTCGATGTAGTCCGCGCT-ACCATGTGACCGGGTGACG 284
|||||
DB 180 TCTGCCCTATTAACTTCGATGTAGTCCGCGCTAACCATGTGACCGGGTGACG 121
|||||
QY 285 GGAATCAGGGTTCGATTCGGAGAGGAGCGCTGAGAAACGGCTACCATCCAGGAAG 344
|||||
DB 120 GGAATCAGGGTTCGATTCGGAGAGGAGCGCTGAGAAACGGCTACCATCCAGGAAG 61
|||||
QY 345 GCAGCAGCGCGCAAAATTACCCACTCCCG 373
|||||
DB 60 GCAGCAGCGCGCAAAATTACCCACTCCCG 32
|||||

Search completed: August 4, 2003, 19:07:44
Job time : 199 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 16:06:37 ; Search time 652 Seconds
(without alignments)
7131.908 Million cell updates/sec

Title: US-10-005-549-1

Perfect score: 2254

Sequence: 1 ctcgctctctctactgg.....tttgaaaaa.....2254

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1554.4	69.0	1666	14	US-10-106-698-181
C 2	432	19.2	432	10	US-09-880-107-3600
C 3	371.4	16.5	870	10	US-09-919-344-14
C 4	371.4	16.5	13076	11	US-09-764-891-8902
C 5	371.4	16.5	13076	11	US-09-764-891-10043
C 6	371.4	16.5	42999	10	US-09-799-462A-17
C 7	371.4	16.5	42999	11	US-09-836-911A-17
C 8	371.4	16.5	42999	13	US-10-125-767-17
C 9	371.4	16.5	42999	14	US-10-151-081-17
C 10	371.4	16.5	42999	14	US-10-287-313-17
C 11	371.4	16.5	42999	14	US-10-219-694-17
C 12	368.2	16.3	1969	10	US-09-880-107-2231
C 13	360	16.0	427	11	US-09-918-995-19104
C 14	352.2	15.6	416	11	US-09-918-995-5393
C 15	349.8	15.5	393	11	US-09-918-995-36028
C 16	333.4	14.8	343	10	US-09-796-692-8493

C 17	333.4	14.8	343	14	US-10-040-862-8493	Sequence 8493, Ap
18	333	14.8	536	14	US-10-106-698-2181	Sequence 2181, Ap
19	331.4	14.7	17384	14	US-10-161-403-118	Sequence 118, Ap
20	331.4	14.7	22118	10	US-09-799-462A-16	Sequence 16, Ap
21	331.4	14.7	22118	11	US-09-815-981-5	Sequence 5, Ap
22	331.4	14.7	22118	11	US-09-836-911A-16	Sequence 16, Ap
23	331.4	14.7	22118	11	US-09-815-979-5	Sequence 5, Ap
24	331.4	14.7	22118	13	US-10-125-767-16	Sequence 16, Ap
25	331.4	14.7	22118	14	US-10-151-081-16	Sequence 16, Ap
26	331.4	14.7	22118	14	US-10-287-313-16	Sequence 16, Ap
27	331.4	14.7	22118	14	US-10-219-694-16	Sequence 16, Ap
28	331.4	14.7	22118	14	US-10-235-119-5	Sequence 5, Ap
29	331.4	14.7	22118	14	US-10-161-403-18	Sequence 18, Ap
C 30	312.6	13.9	696	14	US-10-198-846-9414	Sequence 9414, Ap
C 31	290.8	12.9	488	11	US-09-918-995-17863	Sequence 17863, A
32	281	12.5	281	10	US-09-998-598-2390	Sequence 2390, Ap
33	274.4	12.2	374	13	US-10-079-623-137	Sequence 137, Ap
34	269.4	12.0	316	10	US-09-880-107-194	Sequence 194, Ap
35	262	11.6	510	11	US-09-918-995-16742	Sequence 16742, A
C 36	257.4	11.4	792	13	US-10-027-632-28550	Sequence 28550, A
37	227.4	10.1	357	10	US-09-867-701-2408	Sequence 2408, Ap
C 38	222.8	9.9	822	14	US-10-198-846-9042	Sequence 9042, Ap
C 39	220	9.8	220	10	US-09-878-178-541	Sequence 541, Ap
C 40	220	9.8	220	13	US-10-046-935-541	Sequence 541, Ap
C 41	220	9.8	220	14	US-10-146-502-541	Sequence 541, Ap
C 42	220	9.8	378	10	US-09-799-462A-21	Sequence 21, Ap
43	220	9.8	378	11	US-09-815-981-9	Sequence 9, Ap
44	220	9.8	378	11	US-09-836-911A-21	Sequence 21, Ap
45	220	9.8	378	11	US-09-815-979-9	Sequence 9, Ap

ALIGNMENTS

RESULT 1

US-10-106-698-181/c
; Sequence 181, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 181
; LENGTH: 1666
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (8)..(8)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (1553)..(1553)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-181

Query Match 69.0%; Score 1554.4; DB 14; Length 1666;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1572; Conservative 1; Mismatches 13; Indels 1; Gaps 1;

Qy

661 GAGTCCATCTTCCACCGGCCACGCGG-CTTCCGATTCGCAGGCCCTTCCTGGACGA 719
|||||

Db

1666 GAGTCCATCTTCCACCGGCCACGCGGCTTCCGATTCGCAGGCCCTTCCTGGACGA 1607
|||||

Qy	720	GCTGAGCTGCGAGTTCTGCTGCTGGGGCCCGAGGGGGCCGGGGCGGGGGCGCGCCCGG	7719
Db	1606	GCTGAGCTGCGAGTTCTGCTGCTGGGGCCCGAGGGGGCGGGGGCGGGGGCGCGCCCGG	1547
Qy	780	ACGCGATCTCCCGCCACGGGGGTCGGTGGGATCCTGTCCGATCCTGTCCGATCCACTGCAACAT	839
Db	1546	GACCGCATCTCCCGCCACGGGGTGGTGGGATCCTGTCCGATCCTGTCCGATCCACTGCAACAT	1487
Qy	840	CAGGGATCATACCTGCTGTGCCCCCATCTGGTGGGTGGAGTCTGTAGTGACCCCTAACATT	899
Db	1486	CAGGGATCATACCTGCTGTGCCCCCATCTGGTGGGTGGAGTCTGTAGTGACCCCTAACATT	1427
Qy	900	GGTGTGTCCTTGGAGAGGCTGTGTGGACATAAAGAAAGGGAATACTCTGCTATTGTGACGA	959
Db	1426	GGTGTGTCCTTGGAGAGGCTGTGTGGACATAAAGAAAGGGAATACTCTGCTATTGTGACGA	1367
Qy	960	TCTGAAGAGGATCATCTCCGACCTGTGTAAACTCTATAACTCTCCCTCAGCATCCAGATGT	1019
Db	1366	TCTGAAGAGGATCATCTCCGACCTGTGTAAACTCTATAACTCTCCCTCAGCATCCAGATGT	130
Qy	1020	GGAGATGCTGGATCAACCCCTTGGCAGCAGAGCAGTGCACACAGGAAGACGTGCTTCAGA	1079
Db	1306	GGAGATGCTGGATCAACCCCTTGGCAGCAGAGCAGTGCACACAGGAAGACGTGCTTCAGA	1247
Qy	1080	AGATGAAGATGAGGAGATGCTGTAGGACACAGAAAGACTTTAGATCACTPATGAATGAAAGA	1139
Db	1246	AGATGAAGATGAGGAGATGCTGTAGGACACAGAAAGACTTTAGATCACTPATGAATGAAAGA	1187
Qy	1140	GGAAGGCCAGCTGAGGGCAAGAAATCTGAAGATGATGGCATTTGGAAAAAGAAACTTGGC	1199
Db	1186	GGAAGGCCAGCTGAGGGCAAGAAATCTGAAGATGATGGCATTTGGAAAAAGAAACTTGGC	1127
Qy	1200	CATCCTTAGAAAAATTTAAAAAGAACACAGAGCAAGATTACTTTAAATGGTGCAGTGTCTGG	1259
Db	1126	CATCCTTAGAAAAATTTAAAAAGAACACAGAGCAAGATTACTTTAAATGGTGCAGTGTCTGG	1067
Qy	1260	CTCGGTGCGAGCCACTGACCGCTGTGTGAGGAGCTCAGGGATATATPACCGATCACAGAG	1319
Db	1066	CTCGGTGCGAGCCACTGACCGCTGTGTGAGGAGCTCAGGGATATATPACCGATCACAGAG	1007
Qy	1320	TTTCAAGGGCGGAACATATCAGTCGCAACTCGTGAATGACAGTCTGTATGATGGATGT	1379
Db	1006	TTTCAAGGGCGGAACATATCAGTCGCAACTCGTGAATGACAGTCTGTATGATGGATGT	947
Qy	1380	CAAACTCCTCAAACTTGACAGGACACGGCTTTGCACAACGATCTCCAGATCCTCAAGA	1439
Db	946	CAAACTCCTCAAACTTGACAGGACACGGCTTTGCACAACGATCTCCAGATCCTCAAGA	887
Qy	1440	GAAGAAGGAGCCGACTTCATTCCTAACTTTTCCCTTTAAAGATAAATTTCCCTTTGA	1499
Db	886	GAAGAAGGAGCCGACTTCATTCCTAACTTTTCCCTTTAAAGATAAATTTCCCTTTGA	827
Qy	1500	CCCACCATTTGTCAGGGTTGTCTCCAGTCTCTCTGGAGGTATGTTCTGGGCGGAGG	1559
Db	826	CCCACCATTTGTCAGGGTTGTCTCCAGTCTCTCTGGAGGTATGTTCTGGGCGGAGG	767
Qy	1560	GGCCATCTGCATGGAATTTCTCAACCAAGGGCTGGAGCAGTCCCTACTCCATAGAGTC	1619
Db	766	GGCCATCTGCATGGAATTTCTCAACCAAGGGCTGGAGCAGTCCCTACTCCATAGAGTC	707
Qy	1620	AGTGATCATGCAGATCATGTCACAGACACAGTCTCTCAAGTCCCTTGGTGCAGAT	1739
Db	706	AGTGATCATGCAGATCATGTCACAGACACAGTCTCTCAAGTCCCTTGGTGCAGAT	647
Qy	1680	CAACAAATCTCAATACAGTCTGACAGACACAGTCTCTCAAGTCCCTTGGTGCAGAT	1739
Db	646	CAACAAATCTCAATACAGTCTGACAAAGACACAGTCTCTCAAGTCCCTTGGTGCAGAT	587
Qy	1740	CCAGAAAAAAGCGGTGGTACACACCCCAAGAAAGAGCGCTAAACCCCTGGAGTATCAC	1799
Db	586	CCAGAAAAAAGCGGTGGTACACACCCCAAGAAAGAGCGCTAAACCCCTGGAGTATCAC	527
Qy	1800	CTTCTCCTCCCTCCCGACGCACCTGGACCAATTTACCTTTGAATGCTGTATTTGGATCTC	1859

[illegible]

RESULT 2

```

US-09-880-107-3600/C
; Sequence 3600, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles In Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3600
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Homo sapiens.
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 W79421
US-09-880-107-3600

```

	Query Match	19.2%	Score 432;	DB 10;	Length 432;
	Best Local Similarity	100.0%;	Pred. No. 4.le-104;		
	Matches 432;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1816	GGCACCACTGGACCAATACCTTTGAATGCTGTATTGGATCTCAGCGTCGCCTCTGTGGT	1875		
Dd	432	GGCACCACTGGACCAATACCTTTGAATGCTGTATTGGATCTCAGCGTCGCCTCTGTGGT	373		
QY	1876	TCCCTCCCTCATTTTTCCCTGGACGTGATAGCTCTGCCATTTGCAGGACAATGATGGCTAT	1935		
Dd	372	TCCCTCCCTCATTTTTCCCTGGACGTGATAGCTCTGCCATTTGCAGGACAATGATGGCTAT	313		
QY	1936	TCTAAACGCTAAGGAAAAAAAACACACAGAAGACTGTTTTCAAGTACTCAAGACTGACTTA	1995		


```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5583)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5606)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5953)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (11536)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-891-8902
```

```
Query Match 16.5%; Score 371.4; DB 11; Length 13076;
Best Local Similarity 99.7%; Pred. No. 4e-87;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTCGCTCCTCTCCTTGGATAACTGTGGTAATTCCTAGAGCTAATACATGCCGCGGC 60
Db 3632 CTCGCTCCTCTCCTTGGATAACTGTGGTAATTCCTAGAGCTAATACATGCCGCGGC 3573

Qy 61 GCTGACCCCTTCGCGGGGGGATGCGTGCAATTCATGATCAAGACCAACCCGCTCAGC 120
Db 3572 GCTGACCCCTTCGCGGGGGGATGCGTGCAATTCATGATCAAGACCAACCCGCTCAGC 3513

Qy 121 CCCCTCTCGGCGCGCGGGGGGCGCGCGCGCGCTTGGTGACTCTAGATAACCTC 180
Db 3512 CCCCTCTCGGCGCGCGGGGGGCGCGCGCGCGCTTGGTGACTCTAGATAACCTC 3453

Qy 181 GGGCGGATGCGACGCCCGCGCGCGCGGAGCCATTCGAAGCTCTGCCCTATCAACT 240
Db 3452 GGGCGGATGCGACGCCCGCGCGCGGAGCCATTCGAAGCTCTGCCCTATCAACT 3393

Qy 241 TTCGATGGTAGTCCCGTGCCTACCATGGTGACACCGGGTGACGGGAATCAGGGTTCGA 300
Db 3392 TTCGATGGTAGTCCCGTGCCTACCATGGTGACACCGGGTGACGGGAATCAGGGTTCGA 3333

Qy 301 TTCGGGAGAGGAGCCCTGAGAAACGGCTACACATCAAGGAGGACGCGCGCAAA 360
Db 3332 TTCGGGAGAGGAGCCCTGAGAAACGGCTACACATCAAGGAGGACGCGCGCAAA 3273

Qy 361 TTACCCACTCCCG 373
Db 3272 TTACCCACTCCCG 3260
```

RESULT 5

```
US-09-764-891-10043/c
; Sequence 10043, Application US/09764891
; Publication No. US20030077808a1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10043
; LENGTH: 13076
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1259)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1260)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
```

```
; LOCATION: (1261)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1262)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1263)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1269)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1271)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1274)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1275)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1299)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1311)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1313)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (4089)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5583)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5606)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5953)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (11536)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-891-10043
```

```
Query Match 16.5%; Score 371.4; DB 11; Length 13076;
Best Local Similarity 99.7%; Pred. No. 4e-87;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTCGCTCCTCTCCTTGGATAACTGTGGTAATTCCTAGAGCTAATACATGCCGCGGC 60
Db 3632 CTCGCTCCTCTCCTTGGATAACTGTGGTAATTCCTAGAGCTAATACATGCCGCGGC 3573

Qy 61 GCTGACCCCTTCGCGGGGGGATGCGTGCAATTCATGATCAAGACCAACCCGCTCAGC 120
Db 3572 GCTGACCCCTTCGCGGGGGGATGCGTGCAATTCATGATCAAGACCAACCCGCTCAGC 3513

Qy 121 CCCCTCTCGGCGCGCGGGGGGCGCGCGCGCGCTTGGTGACTCTAGATAACCTC 180
Db 3512 CCCCTCTCGGCGCGCGGGGGGCGCGCGCGCGCTTGGTGACTCTAGATAACCTC 3453

Qy 181 GGGCGGATGCGACGCCCGCGCGCGGAGCCATTCGAAGCTCTGCCCTATCAACT 240
Db 3272 TTACCCACTCCCG 3260
```


TELEFAX: 858-587-5360
TELEX: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/724,693
FILING DATE: 28-NOV-2000
APPLICATION NUMBER: 08/835,682
FILING DATE: 10-APR-1997
APPLICATION NUMBER: 08/695,191
FILING DATE: 07-AUG-1996
APPLICATION NUMBER: 08/682,080
FILING DATE: 15-JUL-1996
APPLICATION NUMBER: 08/629,822
FILING DATE: 10-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24601-402J
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858-450-8403
TELEFAX: 858-587-5360
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 42999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: <Unknown>
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-836-911A-17

Query Match 16.5%; Score 371.4; DB 11; Length 42999;
Best Local Similarity 99.7%; Pred. No. 8.5e-87;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTCGCTCCTCTCCTACTTGGATAACTGTGTAATTTCTAGAGCTAATACATGCCGACGGC 60
Db |||||||
Qy 61 GCTACACCCCTTCGCGGGGGGATGCGTCAATTTATCAGATCAAGACCAACCCGGTCAGC 120
Db |||||||
Qy 3843 GCTACACCCCTTCGCGGGGGGATGCGTCAATTTATCAGATCAAGACCAACCCGGTCAGC 3902
Qy 121 CCCTCTCGGCGCCCGCGGGGGGCGCGCGCGCTTGGTACTCTAGATAACCTC 180
Db |||||||
Qy 3903 CCCTCTCGGCGCCCGCGGGGGGCGCGCGCGCTTGGTACTCTAGATAACCTC 3962
Qy 181 GGGCGGATCGCACGCCCGCCCGTGGCGGCGAGCAGCCCATTCGAACGCTCTGCCCTATCAACT 240
Db |||||||
Qy 3963 GGGCGGATCGCACGCCCGCCCGTGGCGGCGAGCAGCCCATTCGAACGCTCTGCCCTATCAACT 4022
Qy 241 TTCGATGGTAGTCGCGGCTTACCATGGTGACACACCGGTGACCGGGGAATCAGGGTTCGA 300
Db |||||||
Qy 4023 TTCGATGGTAGTCGCGGCTTACCATGGTGACACACCGGTGACCGGGGAATCAGGGTTCGA 4082
Qy 301 TTCGGGAGAGGAGCTTGAGAAACGGCTTACCACATCCAAAGGAAGCGAGCGCGCAAA 360
Db |||||||
Qy 4083 TTCGGGAGAGGAGCTTGAGAAACGGCTTACCACATCCAAAGGAAGCGAGCGCGCAAA 4142
Qy 361 TTACCCACTCCCG 373
Db 4143 TTACCCACTCCCG 4155

RESULT 8:

US-10-125-767-17
Sequence 17, Application US/10125767
Publication No. US20020160410A1
GENERAL INFORMATION:
APPLICANT: Hadlaczky, Gyula
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehrman White & McAuliffe LLP
STREET: 4350 La Jolla Village Drive, 7th Floor
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/125,767

RESULT 9

US-10-151-081-17
Sequence 17, Application US/10151081
Publication No. US20030083293A1
GENERAL INFORMATION:
APPLICANT: Hadlaczky, Gyula

;; Szalay, Aladar
;; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;; NUMBER OF SEQUENCES: 34
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Heller Ehrman White & McAuliffe
;; STREET: 4250 Executive Square, 7th Floor
;; CITY: La Jolla
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 92037
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq Version 1.5
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/151,081
;; FILING DATE: 16-May-2002
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/799,462
;; FILING DATE: HEREWITH 05-MAR-2001
;; APPLICATION NUMBER: 09/724,693
;; FILING DATE: HEREWITH 28-NOV-2000
;; APPLICATION NUMBER: 08/835,682
;; FILING DATE: 10-APR-1997
;; APPLICATION NUMBER: 08/695,191
;; FILING DATE: 07-AUG-1996
;; APPLICATION NUMBER: 08/682,080
;; FILING DATE: 15-JUL-1996
;; APPLICATION NUMBER: 08/629,822
;; FILING DATE: 10-APR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie L
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 24601-402L
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 858-450-8403
;; TELEFAX: 858-587-5360
;; TELEX: <Unknown>
;; INFORMATION FOR SEQ ID NO: 17:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 42999 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Genomic DNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: <Unknown>
;; ORIGINAL SOURCE:
;; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-151-081-17

Query Match 16.5%; Score 371.4; DB 14; Length 42999;
Best local similarity 99.7%; Pred. No. 8.5e-87;
Matches 372; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCTCTCTCTCTGATACTGGTATTTCTAGAGCTAATACATGCGGCGGC 60
DB 3783 CTCGCTCTCTCTCTGATACTGGTATTTCTAGAGCTAATACATGCGGCGGC 3842

QY 61 GGTGACCCCTTCGCGGGGGGATCGTGCATTTATCAGATCAAGACCAACCCGGTCAGC 120
DB 3843 GGTGACCCCTTCGCGGGGGGATCGTGCATTTATCAGATCAAGACCAACCCGGTCAGC 3902

QY 121 CCCTTCGCGGCCCGCGGGGGGCGCGCGCGGCTTTGGTGACTCTAGATAACCTC 180
DB 3903 CCCTTCGCGGCCCGCGGGGGGCGCGCGCGGCTTTGGTGACTCTAGATAACCTC 3962

QY 181 GGGCGCATCGACGCCGCCCGCTGGCGGGGACGACCCATTGCAACCTCTGCCCTATCACT 240
|||||

DB 3963 GGGCGCATCGACGCCGCCCGCTGGCGGGGACGACCCATTCCAAAGTCTGCCCTATCACT 4022
QY 241 TTCGATGTAGTCCCGCTGCTACCATGTGTGACCAACGGGTGACGGGAATCAGGGTTTGA 300
|||||
DB 4023 TTCGATGTAGTCCCGCTGCTACCATGTGTGACCAACGGGTGACGGGAATCAGGGTTGA 4082
QY 301 TTCGGGAGAGGAGCGCTGAGAAACGGGTACCAATCCAAAGAGGAGCGGCGGCAAA 360
|||||
DB 4083 TTCGGGAGAGGAGCGCTGAGAAACGGGTACCAATCCAAAGAGGAGCGGCGGCAAA 4142
QY 361 TTACCCACTCCG 373
|||||
DB 4143 TTACCCACTCCG 4155

RESULT 10
US-10-287-313-17
; Sequence 17, Application US/10287313
; Publication No. US20030101480A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; Szalay, Aladar
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/287,313
; FILING DATE: 01-No. US20030101480A1-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/724,726
; FILING DATE: 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6869-402N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42999 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-287-313-17


```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1969)
; OTHER INFORMATION: n - a or c or g or t
US-09-880-107-2231

Query Match      16.3%; Score 368.2; DB 10; Length 1969;
Best Local Similarity 99.2%; Pred. No. 8.5e-87;
Matches 370; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CTGCTCCTCTCTACTTGGATAACTGTGGTAATCTTAGAGCTAATACATCCGACGGGC 60
      |||||||
Db      228 CTCGCTCCTCTCTACTTGGATAACTGTGGTAATCTTAGAGCTAATACATCCGACGGGC 287
      |||||||

QY      61 GCTGACCCCTTCGCGGGGGGATCGCTGCATTTATCAGATCAAGACCAACCCGGTCAGC 120
      |||||||
Db      288 GCTGACCCCTTCGCGGGGGGATCGCTGCATTTATCAGATCAAAACCCGGTCAGC 347
      |||||||

QY      121 CCCTCTCCGGCCCGCGGGGGCGGGCGCGGCTTTGGTGACTCTAGATAACCTC 180
      |||||||
Db      348 CCCTCTCCGGCCCGCGGGGGCGGGCGCGGCTTTGGTGACTCTAGATAACCTC 407
      |||||||

QY      181 GGGCGGATCGACGCCCCCGCTGCGGCGAGACCCCATTCGAAGCTCTGCCCTATCAACT 240
      |||||||
Db      408 GGGCGGATCGACGCCCCCGCTGCGGCGAGACCCCATTCGAAGCTCTGCCCTATCAACT 467
      |||||||

QY      241 TTGATGGGTAGTCGCGGTGCTTACCATTGGTGACCAACGCGGTACAGGGTTGCGA 300
      |||||||
Db      468 TTGATGGGTAGTCGCGGTGCTTACCATTGGTGACCAACGCGGTACAGGGTTGCGA 527
      |||||||

QY      301 TTCGGAGAGGGAGCCCTGAGAAACGGCTACCAATCCAAAGAGGACGAGCGCGCAAA 360
      |||||||
Db      528 TTCGGAGAGGGAGCCCTGAGAAACGGCTACCAATCCAAAGAGGACGAGCGCGCAAA 587
      |||||||

QY      361 TTACCCACTCCCG 373
      |||||||
Db      588 TTACCCACTCCCG 600
      |||||||

RESULT 13
US-09-918-995-19104
; Sequence 19104, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19104
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-19104

Query Match      16.0%; Score 360; DB 11; Length 427;
Best Local Similarity 100.0%; Pred. No. 4.9e-85;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1995 GGACGTGATAGCTCTGCTATTGCGAGACAATGATGGCTATTCTAAACGCTTAAGAAAAA 1954
      |||||||
Db      40 GGACGTGATAGCTCTGCTATTGCGAGACAATGATGGCTATTCTAAACGCTTAAGAAAAA 99
      |||||||

QY      1955 AAACAAACAGAACTCTTTCAAGTACTCAAGCTACTTACAGACCAACCAACCACTT 2014
      |||||||

```

```
Db      18 TTACAC 13
|||||
RESULT 15
US-09-918-995-36028
; Sequence 36028, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hysq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36028
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36028

Query Match      15.5%; Score 349.8; DB 11; Length 393;
Best Local Similarity 93.1%; Pred. No. 2.3e-82;
Matches 366; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY      1542 GTATGTTCTGGCGGAGGGCCATCTGCATGGAAGTCTTCTACCAAAACAGGCGCTGGAGCAG 1601
|||||
Db      1  GTATGTTCTGGCGGAGGGGTCATCTGCATGGAAGTCTTCTACCAAAACAGGCGCTGGAGCAG 60

QY      1602 TGCCTACTCCATATAGTCAGTGCATGATCATGCAGATCAGTGCACACTGGTGAAGGGGAAAGC 1661
|||||
Db      61  TGCCTACTCCATATAGTCAGAGATCATGCTATCACTGCTCCTGCTGCTGGAAGGGGAAAGC 120

QY      1662 ACGAGTGCAGTTTGGAGGCCAACAAATCTCAATACAGTCTGACAGAGCAGCAGTCCTTA 1721
|||||
Db      121  ACGAGTGCCTGTTGGAGGCCAACAAATGTCATACAGTCTGACAGAGCAGCAGTCCTTA 180

QY      1722 CAAGTCTTGGTGCAGATCCACGAAAAAAGCGCTGGTACACACCCCAAGAGACGG 1781
|||||
Db      181  CAAGTACTTGGTGCAATCCAGAAAAAAGCGCTGTTACACACCCCAAGAGACAG 240

QY      1782 CTAACCCCTGGAGTACACCTTCCTCCCTCCAGGACCACTGGACCAATACCTTTGA 1841
|||||
Db      241  TTAACCCCTGTAGTATTAGCCTTACCTCCCTCCAGGACCACTGGACCAATACCTTTGA 300

QY      1842 ATGCTGTATTGGATCTCACGCTGCTGTGGTTCCCTCCCTCATTTTCTGGACGTG 1901
|||||
Db      301  ATGCTGGATTGGATCTCACGCTGCTGTGGTTCCCTCCCTCATTTTCTGGACGTG 360

QY      1902 ATAGCTCTGCTATTGACAGGACAAATGATGGCTA 1934
|||||
Db      361  ATAGCTCTGCTATTGACAGGACAAATGATGGCTA 393
```

Search completed: August 4, 2003, 21:38:08
Job time : 655 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 12:59:31 ; Search time 5661 Seconds

(without alignments)

9677.140 Million cell updates/sec

Title: US-10-005-549-1

Perfect score: 2254

Sequence: 1 ctcgctctctctctacttg.....tttgaaaaaaaaaaaaa 2254

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 4562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estmu:*

4: em_estov:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rtd:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1217.4	54.0	1408	11 AK009324	AK009324 Mus muscu
2	1083.2	48.1	1201	9 AL540967	AL540967 AL540967
3	1026.6	45.5	1093	13 BU508098	BU508098 AGENCOURT
4	1025.2	45.5	1201	9 AL540968	AL540968 AL540968

c	5	945.6	42.0	1182	9	AL565939	AL565939
c	6	943	41.8	1095	13	BX404169	BX404169
c	7	912.4	40.5	1201	9	AL577525	AL577525
c	8	911.4	40.4	919	13	BX324562	BX324562
c	9	899.8	39.9	1181	9	AL535478	AL535478
c	10	893.4	39.6	1070	12	BM452095	BM452095
c	11	885.8	39.3	1095	13	BX378721	BX378721
c	12	865.4	38.4	895	14	CD107494	CD107494
c	13	841.4	37.3	881	13	BQ212893	BQ212893
c	14	841.4	37.3	1201	9	AL526276	AL526276
c	15	840.2	37.3	1201	9	AL541381	AL541381
c	16	839.6	37.2	1201	9	AL548646	AL548646
c	17	834	37.0	1189	9	AL568878	AL568878
c	18	821.4	36.4	913	13	BU509450	BU509450
c	19	819	36.3	895	13	BX419858	BX419858
c	20	802.2	35.6	1201	9	AL573125	AL573125
c	21	800.2	35.5	1011	12	BM803766	BM803766
c	22	797.4	35.4	852	13	BQ228412	BQ228412
c	23	794	35.2	808	11	AF116721	AF116721
c	24	794	35.2	1238	13	BU515543	BU515543
c	25	789.6	35.0	1201	9	AL554925	AL554925
c	26	775.6	34.4	874	13	BU166850	BU166850
c	27	770.4	34.2	873	13	BQ721866	BQ721866
c	28	768	34.1	949	13	BQ958149	BQ958149
c	29	752.4	33.4	779	12	BQ014708	BQ014708
c	30	747	33.1	758	14	CA449438	CA449438
c	31	733	32.5	735	13	BQ625614	BQ625614
c	32	728.4	32.3	1100	13	BQ278712	BQ278712
c	33	717	31.8	922	13	BQ218761	BQ218761
c	34	714.8	31.7	1201	9	AL526309	AL526309
c	35	714.6	31.7	774	12	BM981520	BM981520
c	36	712.6	31.6	822	12	B1657860	B1657860
c	37	705	31.3	916	13	BX393477	BX393477
c	38	703.8	31.2	837	12	BM011995	BM011995
c	39	703.4	31.2	900	10	BE546982	BE546982
c	40	703	31.2	711	12	BM715207	BM715207
c	41	702.6	31.2	973	10	BQ029803	BQ029803
c	42	697	30.9	755	14	CA453717	CA453717
c	43	691	30.7	743	9	AI207712	AI207712
c	44	686.2	30.4	875	14	CD243557	CD243557
c	45	680.8	30.2	864	12	B1655207	B1655207

ALIGNMENTS

RESULT 1	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
LOCUS	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
DEFINITION	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
ACCESSION	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
VERSION	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
KEYWORDS	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
SOURCE	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
ORGANISM	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
REFERENCE	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
AUTHORS	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
TITLE	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
JOURNAL	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
MEDLINE	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
PUBMED	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
REFERENCE	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
AUTHORS	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
TITLE	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
JOURNAL	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
MEDLINE	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002
PUBMED	AK009324	AK009324	1408 bp	mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310012M18 product:NICE-5 PROTEIN (PRO3094) homolog [Homo sapiens], full insert sequence.	linear	HTC 05-DEC-2002

Db 851 CCTGTGTGGTTCCTCCCTCACTTTTCTGGAGCTGATAGCTCTGCCTATTGCCANNACA 910
 QY 1926 TGATGGCTATTCTTAACGCTTAGGAAAAAACAACACAGAACTGTTTCAAGTACTCAA 1985
 Db 911 TGATGGCTATTCTTAACGCTTAGGAAAAAACAACACAGAACTGTTTCAAGTACTCAA 970
 QY 1986 GACTGACATTAGACACCAACCAACCACTTGTGGAACTTGTCTAGCAGGCACTTCTTATA 2045
 Db 971 GACTGACATTAGCAGCAACCAACCACTTGTCTAGCAGGCACTTCTTATA 1030
 QY 2046 AAGAAACTTTCAGAGCTCTTATATTGCTGAAACTCAGTGTCTCCAGACTAGAGC 2104
 Db 1031 AAAAAAATTTCCAGCTCTTATA-TGCTGGAACTCAGTGTGGCTCAAACTAAACC 1088

RESULT 4

AL540968

LOCUS

DEFINITION AL540968 Homo sapiens PLACENTA 1201 bp mRNA linear EST 12-MAY-2003
 5-PRIME. mRNA sequence.

ACCESSION

AL540968

VERSION

AL540968.2 GI:30544685

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Li.W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished

COMMENT

On Feb 15, 2001 this sequence version replaced gi:12871590.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 78.r For more

information about this cluster, see http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DE004BE02Q1&cluster=78.r. Contact :

Feng liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paraday Avenue Genoscope sequence ID : CS0DE004BE02Q1.

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DE004Y104"

/tissue_type="PLACENTA"

/clone_lib="Homo sapiens PLACENTA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 273 a 303 c 400 g 201 t 24 others

ORIGIN

Query Match

Best Local Similarity 45.5%; Score 1025.2; DB 9; Length 1201;

Matches 1084; Conservative 18; Mismatches 23; Indels 7; Gaps 5;

QY 365 CCACATCCCGGAGGTGGCGGCGGCATCTTGGCGAAGGGGATCAGGAAGTGGCGACC 424

Db 71 CCTGTCCGGAGGTGGCGGCGGCATCTTGGCGAAGGGGATCAGGAAGTGGCGACC 130

QY 425 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 484

Db 131 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 187

QY 485 CCCGCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 544

Db 188 CCCCG 247
 QY 545 AGCAGCCGGGCG 604
 Db 248 AGCAGCCGGGCG 307
 QY 605 GCCCAGGG 664
 Db 308 GCCCAGGG 367
 QY 665 CCATCTTCCACCGGGGCG 724
 Db 368 YCATCTTCCACCGGGGCG 427
 QY 725 GCTCGGAGTTCTCTGCTGGGCGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 784
 Db 428 GCTCGGAGTTCTCTGCTGGGCGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 487
 QY 785 ATCTCCCCCAGCGGGGCTGGTGGGATCTCTGGGATCTCTGGGATCTCTGGGATCTCTGG 844
 Db 488 ATCTCCCCCAGCGGGGCTGGTGGGATCTCTGGGATCTCTGGGATCTCTGGGATCTCTGG 546
 QY 845 AGTCATACCTGCTGTGCCCGCCATCTGGTGGGATCTCTGGGATCTCTGGGATCTCTGGG 904
 Db 547 AGTCATACCTGCTGTGCCCGCCATCTGGTGGGATCTCTGGGATCTCTGGGATCTCTGGG 606
 QY 905 CTGTCTTGGAGAGCTGGTGGACATAAAGAAAGGAATACTCTCTTATTTGACGATCTGA 964
 Db 607 CTGTCTTGGAGAGCTGGTGGACATAAAGAAAGGAATACTCTCTTATTTGACGATCTGA 666
 QY 965 AGAGGATCATCTCCGACCTGTGTAACCTCTATAACCTCTCCCTCAGCATCTGAGATG 1024
 Db 667 AGAGGATCATCTCCGACCTGTGTAACCTCTATAACCTCTCCCTCAGCATCTGAGATG 726
 QY 1025 TGCTGGATCAACCTTCCAGCAGCAGTGCACACAGGAGAGAGTCTCTCAGAAATG 1084
 Db 727 TGCTGGATCAACCTTCCAGCAGCAGTGCACACAGGAGAGAGTCTCTCAGAAATG 786
 QY 1085 AAGATGAGGAGATGCTTGGAGACAGAGAACTTAGATCACTATGAATGAAGAGGAAG 1144
 Db 787 AAGATGAGGAGATGCTTGGAGACAGAGAACTTAGATCACTATGAATGAAGAGGAAG 846
 QY 1145 AGCCAGCTGAGGGCAAGAAATCTGAAGATGATGCAATGGAAGAAATCTGGCCATCC 1204
 Db 847 AGCCAGCTGAGGGCAAGAAATCTGAAGATGATGCAATGGAAGAAATCTGGCCATCC 906
 QY 1205 TAGAGAAATTTAAAGAACAGAGCAAGATTAATTAATGTTGAGTCTCTGGCTCGG 1264
 Db 907 TAGAGAAATTTAAAGAACAGAGCAAGATTAATTAATGTTGAGTCTCTGGCTCGG 966
 QY 1265 TGCAGGCACTGACCGGCTGATGAAGAGGCTCAGGGATATATACCGATCAGAGTTTCA 1324
 Db 967 TGCAGGCACTGACCGGCTGATGAAGAGGCTCAGGGATATATACCGATCAGAGTTTCA 1026
 QY 1325 AAGCGGAAATATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1384
 Db 1027 AAGCGGAAATATGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1085
 QY 1385 TCCTCAAGTTGACCAAGAGAGCGCTTGGACACAGATCTCCAGATCTCAAGAGAAAG 1444
 Db 1086 TCCTCAAGTTGACCAAGAGAGCGCTTGGACACAGATCTCCAGATCTCAAGAGAAAG 1145
 QY 1445 AAGCGGCACTTCTTACTTAACTTTTCTTAAAGAAATCTTTCCCTTT 1496
 Db 1146 AAG--GGCGGCTTTTCTTAAAGAAATCTTTCCCTTT 1195

RESULT 5

AL565939/c

LOCUS

DEFINITION

AL565939 Homo sapiens FETAL, BRAIN Homo sapiens cDNA clone

CS0DF009Y108 3-PRIME, mRNA sequence.

ACCESSION

AL565939

1182 bp mRNA linear EST 12-MAY-2003

[illegible]

[illegible]


```
Db 975 CAAGATCATTAAATGKAGTGTCTGGTCGGTCAGCCCACTGACCGCTGATGAAG 916
Qy 1291 GAGCTCAGGATATATACCGATCACAGAGTTTCAAGCGCGAACTATGAGTCGAAGTC 1350
Db 915 GRCCTCAGGATATATACCGATCACAGAGTTTCAAGCGCGAACTAGGAGTCGAAGTC 856
Qy 1351 GTGAATGACAGCTGTATGATGGAATGTCAAACTCTCAAGTTGACCGAGCAGCGCT 1410
Db 855 GTGAATGACAGCTGTATGATGGAATGTCAAACTCTCAAGTTGACCGAGCAGCGCT 796
Qy 1411 TTGCACACAGATCTCCAGATCTCAAGAGAGAAAGAGCGGCACTTCACTACTTAAC 1470
Db 795 TTGCACACAGATCTCCAGATCTCAAGAGAGAAAGAGCGGCACTTCACTACTTAAC 736
Qy 1471 TTTTCCCTTAAAGATAACTTTCCCTTTGACCCACCATTTGTACGGTGTGTCTCCAGTC 1530
Db 735 TTTTCCCTTAAAGATAACTTTCCCTTTGACCCACCATTTGTACGGTGTGTCTCCAGTC 676
Qy 1531 CTCTCTGGAGGATATGTTCTGGCGGAGGGGCCATCTGCATGGAACCTTCTACCAACAG 1590
Db 675 CTCTCTGGAGGATATGTTCTGGCGGAGGGGCCATCTGCATGGAACCTTCTACCAACAG 616
Qy 1591 GGCTGGAGCAGTCTACTCTCATAGATGATGATGATGATGATGATGATGATGATGATG 1650
Db 615 GGCTGGAGCAGTCTACTCTCATAGATGATGATGATGATGATGATGATGATGATGATG 556
Qy 1651 AAGGGGAAGCAGCAGTGTGAGTGTGGAGCCCAACAAATCTCAATACAGTCTGACAGAGCA 1710
Db 555 AAGGGGAAGCAGCAGTGTGAGTGTGGAGCCCAACAAATCTCAATACAGTCTGACAGAGCA 496
Qy 1711 CAGCAGTCTCAAGTCTCTGGTGAGATCCACGAAACAAAGGCTGGTACACACCCCA 1770
Db 495 CAGCAGTCTCAAGTCTCTGGTGAGATCCACGAAACAAAGGCTGGTACACACCCCA 436
Qy 1771 AAAGAGAGCCTTAACCTGAGTATACCTCTCCCTCCCGAGCAGCAGTGGAGCA 1830
Db 435 AAAGAGAGCCTTAACCTGAGTATACCTCTCCCTCCCGAGCAGCAGTGGAGCA 376
Qy 1831 ATTACCTTTGAATGCTGATTTGGATCTCAGCGTCCCTCTGTGTGTCCTCCCTCATTTT 1890
Db 375 ATTACCTTTGAATGCTGATTTGGATCTCAGCGTCCCTCTGTGTGTCCTCCCTCATTTT 316
Qy 1891 TCCTGGAGCTGATGCTGCTGCTATGTCAGGACAAATGATGCTATTTAAAGCTT-AGG 1949
Db 315 TCCTGGAGCTGATGCTGCTGCTATGTCAGGACAAATGATGCTATTTAAAGCTTAAAGA 256
Qy 1950 AAAAAACAAACACAGAACTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACC 2009
Db 255 AAAAAACAAACACAGAACTGTTTCAAGTACTCAAGACTGACTTACAGACCAACCAACC 196
Qy 2010 ACCTTGTGGAACTTCTGCTAGCAGGCACTTCTTATAAAGAACTTTGAGCCTCTTAT 2069
Db 195 ACCTNGCTGGAACCTTCTGCTAGCAGGCACTTCTTATAAAGAACTTTGAGCCTCTTAT 136
Qy 2070 ATTGCTGGAACCTCAGCTGTGCTCAGACTAGAGCCTCTTACCTATGCTATGATTTT 2129
Db 135 ATNGCTGGAACCTCAGCTGTGCTCAGACTAGAGCCTCTTACCTATGCTATGATTTT 76
Qy 2130 AATTTATTTCTCTTATTTATGATGATGATGATGATGATGATGATGATGATGATG 2166
Db 75 AATTTATTTCTCTTATTTATGATGATGATGATGATGATGATGATGATGATGATG 39
```

RESULT 8

```

BX324562/c
LOCUS BX324562 919 bp mRNA linear EST 01-MAY-2003
DEFINITION BX324562 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
          CDNA clone CS0DC025YA19 3-PRIME, mRNA sequence.
ACCESSION BX324562
VERSION   BX324562.1 GI:30309380
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
```

```

ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 919)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
          Library was constructed by Life Technologies, a division of
          Invitrogen. This sequence belongs to sequence cluster 78.r For more
          information about this cluster, see http://www.genoscope.cns.fr/
          cgi-bin/cluster.cgi?seq=CS0AC025AA10NP1&cluster=78.r. Contact :
          Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
          Paraday Avenue Genoscope sequence ID : CS0AC025AA10NP1.
FEATURES             Location/Qualifiers
     source            1..919
     -organism="Homo sapiens"
     -mol_type="mRNA"
     -db_xref="taxon:9606"
     -clone="CS0DC025YA19"
     -tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
     -clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
     -note="1st strand cDNA was primed with a NotI-oligo(dT)
     primer. Five prime end enriched, double-strand cDNA was
     digested with Not I and cloned into the Not I and EcoR V
     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT  239 a  185 c  229 g  255 t  11 others
ORIGIN
Query Match      40.4%; Score 911.4; DB 13; Length 919;
Best Local Similarity 98.8%; Pred. No. 3.7e-134;
Matches 908; Conservative 10; Mismatches 1; Indels 0; Gaps 0;
Qy 1304 TATACCGATCACAGAGTTTCAAGCGGAAACTATGAGTCGAACTCGTGAATGACAGTC 1363
Db 919 TATACCGATCACAGAGTTTCAAGCGGAAACTATGAGTCGAACTCGTGAATGACAGTC 860
Qy 1364 TGATGATGGAATGTCAAACTCTCTCAAGTTGACCGAGCAGCGCTTTGCACAAAGATC 1423
Db 859 TGATGATGGAATGTCAAACTCTCTCAAGTTGACCGAGCAGCGCTTTGCACAAAGATC 800
Qy 1424 TCAGATCTCTCAAGAGAAAGAGGAGCGGACTTCTTACTTAACTTTTCTTTAAAG 1483
Db 799 TCAGATCTCTCAAGAGAAAGAGGAGCGGACTTCTTACTTAACTTTTCTTTAAAG 740
Qy 1484 ATAACTTTCCCTTTGACCCACCACTTTGTGTCAGGTTGTGTCTCCAGTCTCTGAGGGT 1543
Db 739 ATAACTTTCCCTTTGACCCACCACTTTGTGTCAGGTTGTGTCTCCAGTCTCTGAGGGT 680
Qy 1544 ATGTTCTGGCGGAGGGGCCATCTGATGAACTTCTCACCACAGGCGCTGAGCAGTG 1603
Db 679 ATGTTCTGGCGGAGGGGCCATCTGATGAACTTCTCACCACAGGCGCTGAGCAGTG 620
Qy 1604 CCTACTCCATAGAGTCACTGATCATGTCAGATCATGTCAGTCTGAGGAGGAAAGAC 1663
Db 619 CCTACTCCATAGAGTCACTGATCATGTCAGTCTGAGGAGGAAAGAC 560
Qy 1664 GAGTCGAGTTGGAGCCCAACAAATCTCAATACAGTCTGACAGCAGCAGCAGTCTCTACA 1723
Db 559 GAGTCGAGTTGGAGCCCAACAAATCTCAATACAGTCTGACAGCAGCAGCAGTCTCTACA 500
Qy 1724 AGTCTTGTGTCAGATCCAGCAAAAAAAGCGCTGGTGTACACACCCCAAGAGAGCGCT 1783
Db 499 AGTCTTGTGTCAGATCCAGCAAAAAAAGCGCTGGTGTACACACCCCAAGAGAGCGCT 440
Qy 1784 AACCTTGGAGTATCACCTTCTCCCTCCCGAGGACCACTGGACCAATTTACCTTTGAAT 1843
Db 439 AACCTTGGAGTATCACCTTCTCCCTCCCGAGGACCACTGGACCAATTTACCTTTGAAT 380
```



```
Qy 1428 GATCCTCAAGAGAGAGAGAGCGGACCTTCTACTTAACTTTTCCCTTTAAAGATAA 1487
| | | | | : : : | | | | | : | | | | | : | | | | |
Db 1072 GKGAARAAAAAGWKGRTTCTGTTTATTARKTTTGTAAATAARTTGKTTAVGGGA 1131
| | | | | : : : | | | | | : | | | | | : | | | | |
Qy 1488 CTTTCCCTTGGACCCACCAATTT 1509
| | | | | : : : | | | | | : | | | | | : | | | | |
Db 1132 TTTWAGCTTTTTSRTYSTTTT 1153
| | | | | : : : | | | | | : | | | | | : | | | | |

RESULT 10
BM452095 1070 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6386309 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5526427
DEFINITION 5', mRNA sequence.
ACCESSION BM452095
VERSION BM452095.1 GI:18501135
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1070)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ANCC/DCTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12200 Row: a Column: 20
High quality sequence stop: 653.
Location/Qualifiers
1..1070
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5526427"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/note="Organ: skin; Vector: pCMV-Sport6; Site:1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 312 a 270 c 238 g 250 t
ORIGIN

Query Match 39.6%; Score 893.4; DB 12; Length 1070;
Best Local Similarity 96.9%; Pred. No. 2.4e-131;
Matches 953; Conservative 0; Mismatches 26; Indels 4; Gaps 4;

Qy 1089 TGAGGAGATGCTGAGGACACAGAGAGCTTATGATCACTATGAAATGAAGAGGAGAGCC 1148
| | | | | : : : | | | | | : | | | | | : | | | | |
Db 1 TGAGGAGATGCTGAGGACACAGAGAGCTTATGATCACTATGAAATGAAGAGGAGAGCC 60
| | | | | : : : | | | | | : | | | | | : | | | | |

Qy 1149 AGCTGAGGGCAAGAAATCTGAAGATGATGGCATTGGAAAGAAAACCTTGGCCATCTCTAGA 1208
| | | | | : : : | | | | | : | | | | | : | | | | |
Db 61 AGCTGAGGGCAAGAAATCTGAAGATGATGGCATTGGAAAGAAAACCTTGGCCATCTCTAGA 120
| | | | | : : : | | | | | : | | | | | : | | | | |

Qy 1209 GAAATTTAAAGAACACAGAGCAAGATTACTTAATGGTCAGTCTGGCTCGGTGCA 1268
| | | | | : : : | | | | | : | | | | | : | | | | |
Db 121 GAAATTTAAAGAACACAGAGCAAGATTACTTAATGGTCAGTCTGGCTCGGTGCA 180
| | | | | : : : | | | | | : | | | | | : | | | | |

Qy 1269 GCCACTGACCGGCTGATGAGGAGCTCAGGATATATACCATCACAGAGTTTCAAGG 1328
| | | | | : : : | | | | | : | | | | | : | | | | |
Db 181 GCCACTGACCGGCTGATGAGGAGCTCAGGATATATACCATCACAGAGTTTCAAGG 240
| | | | | : : : | | | | | : | | | | | : | | | | |

Qy 1329 CGGAACATATCGACTCGAAGCTCGTGAATGACAGCTGTGATGATTGGAATGTCAAACTCCT 1388
| | | | | : : : | | | | | : | | | | | : | | | | |
```

RESULT 11

```
BX378721
LOCUS BX378721 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1024YM15 5-PRIME, mRNA sequence.
ACCESSION BX378721
VERSION BX378721.1 GI:30460150
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1095)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
```

FEATURES

Query Match 39.3%; Score 885.8; DB 13; Length 1095;
Best Local Similarity 93.9%; Pred. NO. 3.8e-130;
Matches 978; Conservative 15; Mismatches 39; Indels 9; Gaps 7;

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 13:35:23 ; Search time 83 Seconds
(without alignments)
807.019 Million cell updates/sec

Title: US-10-005-549-2

Perfect score: 2229

Sequence: 1 MQQPQPGQQPGPQQQLGG.....KSLVQIHEKNGWYPPPKEDG 422

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2229	100.0	422	23 AAE24493	Human RAT1d6 (reg
2	1863	83.6	374	22 AAG73684	Human colon cancer
3	1509	67.7	291	20 AAY48312	Human prostate can
4	1300.5	58.3	340	23 ABB97360	Novel human protei
5	1293	58.0	245	23 AAE24494	Human RAT1d6 tran
6	1266	56.8	263	21 AAB41773	Human ORFX ORF1537
7	1120.5	50.3	308	21 AAY73387	HTRM clone 3340290
8	1020.5	45.8	397	22 ABB59516	Drosophila melanog
9	834	37.4	164	23 AAE24524	Human RAT1d6 ubiq

ALIGNMENTS

RESULT 1

AAE24493
ID AAE24493 standard; Protein; 422 AA.

XX AAE24493;

AC AAE24493;

DT 04-OCR-2002 (first entry)

DE Human RAT1d6 (regulated in activated T-lymphocyte ld6) protein.

Human; ubiquitin conjugating enzyme; UBC; RAT1d6; immune disorder; regulated in activated T-lymphocyte ld6; neuronal disorder; cancer; tumour; lymphoproliferative; cancer; adenocarcinoma; leukaemia; myeloma; sarcoma; neurodegenerative; inflammatory; rheumatoid arthritis; asthma; multiple sclerosis; psoriasis; neuronal; Alzheimer's disease; dementia; depression; epilepsy; acquired immuno deficiency syndrome; allergy; AIDS; anaemia; atopic dermatitis; diabetes mellitus; dermatological; myocardial infarction; renal tubular acidosis; gonadal dysgenesis; dysplasia; cataract; cytostatic; neuroprotective; nontropic; anti-HIV; anticonvulsant; antiinflammatory; Cushing's syndrome; cardiact;

XX Homo sapiens.

XX Key Location/Qualifiers

XX Domain 89..333

XX /note= "Transmembrane domain"

XX WO200236741-A2.

XX 10-MAY-2002.

XX

10	808	36.2	217	22	ABB67021	Drosophila melanog
11	440.5	19.8	306	22	ABB60719	Drosophila melanog
12	440.5	19.8	306	22	ABB60722	Drosophila melanog
13	304	13.6	747	22	ABG11969	Novel human diagn
14	260	11.7	400	22	ABG28233	Novel human diagn
15	187.5	8.4	1909	22	ABG15534	Novel human diagn
16	163	7.3	452	22	ABG20796	Novel human diagn
17	154	6.9	59	22	ABG03538	Novel human diagn
18	140.5	6.3	46	22	ABG54524	Human liver peptid
19	140.5	6.3	46	22	ABB39508	Peptide #7014 enco
20	140.5	6.3	46	22	AAM60200	Human brain expres
21	140.5	6.3	46	22	AAM72820	Human bone marrow
22	140.5	6.3	46	22	AAM33050	Peptide #7087 enco
23	140.5	6.3	46	22	ABG42649	Human peptid enco
24	135.5	6.1	295	23	AAE28245	Yeast Cdc34 protei
25	135	6.1	199	22	ABE63741	Drosophila melanog
26	134.5	6.0	44	22	ABG03535	Novel human diagn
27	134.5	6.0	223	24	ABJ26294	Aspergillus fumiga
28	133	6.0	238	24	ABJ25694	Aspergillus fumiga
29	131	5.9	244	23	ABP73432	Candida albicans e
30	128.5	5.8	190	22	ABG63097	Drosophila melanog
31	128	5.7	226	23	ABP41187	Human ovarian anti
32	127.5	5.7	207	21	ABA43031	Human ORFX ORF2795
33	127.5	5.7	207	24	ABP71416	Human UBE2 related
34	125	5.6	81	22	AAO02768	Human polypeptide
35	125	5.6	1235	23	ABB77435	Human tumour marke
36	123	5.5	463	21	ABA41079	Human ORFX ORF843
37	123	5.5	914	23	ABP43569	FLJ10378 fls clone
38	123	5.5	963	24	ABJ19401	Human intracellular
39	122.5	5.5	177	18	AAW31278	Clam cyclin-select
40	122.5	5.5	1048	21	AAG42118	Arabidopsis thalia
41	122	5.5	477	22	ABBS9034	Drosophila melanog
42	121	5.4	542	22	ABG20670	Novel human diagn
43	120.5	5.4	716	14	AA307030	B. burgdorferi 79
44	120.5	5.4	924	23	ABBS4116	Lactococcus lactis
45	120.5	5.4	1004	22	ABG19122	C glutamicum prote

PF 29-OCT-2001; 2001WO-US46559.
 XX
 PR 30-OCT-2000; 2000US-244688P.
 PR 30-JUL-2001; 2001US-308706P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Bowen MA, Wu Y, Yang W, Finger JN;
 XX
 DR WPI; 2002-479758/51.
 DR N-PSDB; RAD39359.
 XX
 XX Novel ubiquitin conjugating enzyme polypeptide isolated from activated
 PT human T cell, for screening modulators useful for treating cancer,
 PT immune disorder, lymphoproliferative disorder, neurodegenerative
 PT disorder
 XX
 PS Claim 6; Fig 3; 169pp; English.
 XX
 CC The invention relates to a novel ubiquitin conjugating enzyme (UBC)
 CC homologue, RAT1d6 (regulated in activated T-lymphocytes 1d6) and its
 CC corresponding nucleic acid. The invention also relates to methods for
 CC treating, diagnosing, preventing and screening for disorders related
 CC to the expression of RAT1d6. UBC is useful for screening for candidate
 CC compounds capable of binding to and/or modulating its activity. UBC is
 CC useful for treating an immune or neuronal disorder in a mammal. The
 CC method is useful for treating a cancer or tumour. It is useful for
 CC suppressing the immune response in a subject requiring the suppression.
 CC It is also useful for treating lymphoproliferative disorder, cancer e.g.
 CC adenocarcinoma, leukaemia, myeloma, sarcoma, etc, neurodegenerative
 CC disorder, inflammatory disorders e.g. rheumatoid arthritis, asthma,
 CC multiple sclerosis, psoriasis, etc, neuronal disorders e.g. Alzheimer's
 CC disease, dementia, depression, epilepsy, etc, immune disorder or immune
 CC related disorders such as acquired immuno deficiency syndrome (AIDS),
 CC allergy, anaemia, atopic dermatitis, diabetes mellitus, myocardial
 CC infarction, etc, developmental disorders e.g. Cushing's syndrome, renal
 CC tubular acidosis, gonadal dysgenesis, dysplasia, cataract, etc. The
 CC present sequence is human RAT1d6 protein.
 XX
 SQ Sequence 422 AA;
 Query Match 100.0%; Score 2229; DB 23; Length 422;
 Best Local Similarity 100.0%; Pred. No. 7.7e-181;
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQQPOGQOQPGGQQLGGGGAAPGAGGGGPGGPGCLRRRLKLLSFHGRHFR 60
 DB 1 MQQPOGQOQPGGQQLGGGGAAPGAGGGGPGGPGCLRRRLKLLSFHGRHFR 60
 QY 61 IASACDLSCEFLLAGAGGAGAGAPCPHLPGRSGVPGDPVRIHCNITESYPAPVPIWS 120
 DB 61 IASACDLSCEFLLAGAGGAGAGAPCPHLPGRSGVPGDPVRIHCNITESYPAPVPIWS 120
 QY 121 VESDDPNLAVALERLVDIKKNTLLQHLKRIISDLCKLYNLPQHPDVMELDQPLPAEQ 180
 DB 121 VESDDPNLAVALERLVDIKKNTLLQHLKRIISDLCKLYNLPQHPDVMELDQPLPAEQ 180
 QY 181 TQEDVSEDEDEPEDEDLDHYEMKEERPAEGKSEDDGIGKENLAILEKIKKNORQD 240
 DB 181 TQEDVSEDEDEPEDEDLDHYEMKEERPAEGKSEDDGIGKENLAILEKIKKNORQD 240
 QY 241 YLNGAVSGSVQATDRLMKELRDYRSQSFSGNGYAVELVNDSDLYDNVKKLVKDQDSALH 300
 DB 241 YLNGAVSGSVQATDRLMKELRDYRSQSFSGNGYAVELVNDSDLYDNVKKLVKDQDSALH 300
 QY 301 NDLQILKEGADIFLLNFSKDNFPDPPFVRVSVPLSGGYVLGGGATCMELLTKQGW 360
 DB 301 NDLQILKEGADIFLLNFSKDNFPDPPFVRVSVPLSGGYVLGGGATCMELLTKQGW 360
 QY 361 SSAYSIESVIMQISATLVKGRARVQFGANKSQYSILTRAQOQSKYSLVQIHEKNQWYTPPKE 420
 DB 361 SSAYSIESVIMQISATLVKGRARVQFGANKSQYSILTRAQOQSKYSLVQIHEKNQWYTPPKE 420

QY 421 DG 422
 DB 421 DG 422
 RESULT 2
 AAG73684
 ID AAG73684 standard; Protein; 374 AA.
 XX
 XX AAG73684;
 XX
 XX 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:4448.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;
 PI WPI; 2001-235357/24.
 XX
 DR N-PSDB; AAH33115.
 DR
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 11; Page 6262-6264; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAG77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 374 AA;
 Query Match 83.6%; Score 1863; DB 22; Length 374;
 Best Local Similarity 98.4%; Pred. No. 8.6e-150;
 Matches 359; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 59 FRIASACDLSCEFLLAGAGGAGAGAPCPHLPGRSGVPGDPVRIHCNITESYPAPVPI 118
 DB 11 FRIASACDLSCEFLLAGAGGAGAGAPCPASPTPGDPVRIHCNITESYPAPVPI 70
 QY 119 WSVESDDPNLAVALERLVDIKKNTLLQHLKRIISDLCKLYNLPQHPDVMELDQPLPAE 178

Db 71 WSVESDDNLAALVRLVDIKKNTLLQLHLKRIISDLCKLYNLPQHPDVEMLDQPLPAE 130
QY 179 OCTQEDVSSDEDEMPEDTEDLDHYEMKEEPAEGKSKSDDGIGKENLAILEIKKKNOR 238
Db 131 OCTQEDVSSDEDEMPEDTEDLDHYEMKEEPAEGKSKSDDGIGKENLAILEIKKKNOR 190
QY 239 QDYLNAGVSGVQATDRLMKELRDYRSQSFKGGNYAVELVNDSLYDNNVLLKVDQDSA 298
Db 191 QDYLNAGVSGVQATDRLMKELRDYRSQSFKGGNYAVELVNDSLYDNNVLLKVDQDSA 250
QY 299 LNDLQILKEKEGADFIILNFSKDNFPDPFVRVSPVLSGGYVLGGGAICMELLTKQ 358
Db 251 LNDLQILKEKEGADFIILNFSKDNFPDPFVRVSPVLSGGYVLGGGAICMELLTKQ 310
QY 359 GWSAYSIESYIMQISATLVKGRVQFGANKSQYSLTRAQSQYKSLVQIHEKNGWYTPP 418
Db 311 GWSAYSIESYIMQISATLVKGRVQFGANKSQYSLTRAQSQYKSLVQIHEKNGWYTPP 370
QY 419 KEDG 422
Db 371 KEDG 374

RESULT 3
ID AAY48312 standard; Protein; 291 AA.
AC AAY48312;

DT 08-DEC-1999 (first entry)
XX Human prostate cancer-associated protein 9.
DE Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
KW cancer; tissue specificity; human.
XX Homo sapiens.

XX DE19811194-A1.
XX 16-SEP-1999.
XX 10-MAR-1998; 98DE-1011194.
XX 10-MAR-1998; 98DE-1011194.
XX (META-) METAGEN GES. GENOMFORSCHUNG MBH.
XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX WPI; 1999-519629/44.
XX N-PSDB; AAZ33482.

XX New nucleic acid expressed at high level in normal prostatic tissue and
XX encoded polypeptides, used to treat cancer and screen for therapeutic
XX agents
XX Claim 22; 128; 194pp; German.

XX This invention describes novel nucleic acid sequences (A) that are
XX expressed at high level in normal prostatic tissue. Polypeptides (I)
XX encoded by (A) are used: (a) for identifying agents for treatment of
XX prostatic cancer and (b) for therapy of prostate cancer, optionally
XX where expressed by gene therapy methods. (A) is also used to isolate
XX full-length genes (for gene therapy) and for recombinant production of
XX (I), which can be used to raise specific antibodies. (A) are identified
XX by assembly of ESTs (expressed sequence tags) before these are analyzed
XX for expression pattern (tissue specificity). This approach eliminates
XX many of the false results, as regards tissue specificity, associated
XX with known methods that use single (usually short) ESTs. AAY48304-Y48456
XX represent peptides encoded by the expressed sequence tags described in
XX the method of the invention.

SQ Sequence 291 AA;

Query Match 67.7%; Score 1509; DB 20; Length 291;
Best Local Similarity 100.0%; Pred. No. 7.6e-120;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 132 LERLVDIKKNTLLQLHLKRIISDLCKLYNLPQHPDVEMLDQPLPAEQCTQEDVSSDED 191
Db 1 LERLVDIKKNTLLQLHLKRIISDLCKLYNLPQHPDVEMLDQPLPAEQCTQEDVSSDED 60
QY 192 EEMPEDETDLDHYEMKEEPAEGKSKSDDGIGKENLAILEIKKKNQDYLNAGVSGVQ 251
Db 61 EEMPEDETDLDHYEMKEEPAEGKSKSDDGIGKENLAILEIKKKNQDYLNAGVSGVQ 120
QY 252 ATDRLMKELRDYRSQSFKGGNYAVELVNDSLYDNNVLLKVDQDSALHNDLQILKEK 311
Db 121 ATDRLMKELRDYRSQSFKGGNYAVELVNDSLYDNNVLLKVDQDSALHNDLQILKEK 180
QY 312 ADFILLNFSKDNFPDPFVRVSPVLSGGYVLGGGAICMELLTKQGSAYSIESYIM 371
Db 181 ADFILLNFSKDNFPDPFVRVSPVLSGGYVLGGGAICMELLTKQGSAYSIESYIM 240
QY 372 QISATLVKGRVQFGANKSQYSLTRAQSQYKSLVQIHEKNGWYTPPKEDG 422
Db 241 QISATLVKGRVQFGANKSQYSLTRAQSQYKSLVQIHEKNGWYTPPKEDG 291

RESULT 4
ID ABB97360 standard; Protein; 340 AA.

XX ABB97360;
XX 27-JUN-2002 (first entry)
XX Novel human protein SEQ ID NO: 628.
DE Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.

XX Homo sapiens.

XX WO200222660-A2.

XX 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US26015.

XX 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-292408/33.

XX N-PSDB; ABN32546.

XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -

XX Example 2; SEQ ID NO 628; 509pp; English.

XX The present invention provides the protein and coding sequences of 414
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat
XX stroke and cancer, to screen for drugs, to treat inflammatory conditions

CC	e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC	Parkinson's disease. The present sequence is a protein of the invention.
XX	
SQ	Sequence 340 AA;
	Query Match 58.3%; Score 1300.5; DB 23; Length 340;
	Best Local Similarity 67.8%; Pred. No. 5,1e-102;
	Matches 261; Conservative 30; Mismatches 41; Indels 53; Gaps 5;
QY	41 LRRELKLESIFRUGHERRPIASACLDLSCFLLAGAGAGAGAAPGPH-LPPRGSGVPG 99
Db	6_LKAEFLASIFDKNHERFRIVSWKLDLHCQLVPOQG-----SPHSLPP----- 51
QY	100 DPVRIHCNITESYPAPPPISVSDDDPNLAANLRLVDIKKGNLTLLQHLKRIISDLCKL 159
Db	52 -PLTLHCNITESYPPSSPIFVDSDEPNLTSVLERLEDTKNNL----- 94
QY	160 YNLPQHPDVENLDQPLPAEOCTQEDYSS--EDDEEMPEDTEDLDHYEMKEEPEAGKKS 217
Db	95 -----NGTTEVTSEEEEEEMAIEDIEDLDHYEMKEEPIISGKS 135
QY	218 EDDGIGKENLAILEKIKKNORQDYLVNGAVSGSVQATDRLMKELRDIVRSOFKGGNYAVE 277
Db	136 EDGIEKENLAILEKIKRKTQRQDHLNGAVSGSVQASDRLMKELRDIVRSQIKGTIYSVE 195
QY	278 LVNDSLYDNVKKLVKVDQDSALHNDLQILKEGADFIILNFSFKONFPDPDPFVRYVSP 337
Db	196 LINDSLDYHVKLQKVPDPSPLHSDLIQLEKEGIEYILLNFSFKDNFPDPDPFVRYVLP 255
QY	338 VLSGGYVLGGATCMELLTQGWSSAYSIESVIMQISATLVKGKARVQFGANKSQYSLTR 397
Db	256 VLSGGYVLGGALCMELLTQGWSSAYSIESVIMQINATLVKGKARVQFGANKQYNLAR 315
QY	398 AQQSYKSLVQIHEKNGWYTPPKEDG 422
Db	316 AQQSYNSIVQIHEKNGWYTPPKEDG 340
RESULT 5	
AAE24494	
ID	AAE24494 standard; Protein; 245 AA.
XX	
AC	AAE24494;
XX	
DT	04-OCT-2002 (first entry)
XX	
DE	Human RATL1d6 transmembrane domain.
XX	
KW	Human; ubiquitin conjugating enzyme; UBC; RATL1d6; Immune disorder;
KW	regulated in activated T-lymphocyte 1d6; neuronal disorder; cancer;
KW	tumour; lymphoproliferative; cancer; adenocarcinoma; leukaemia; myeloma;
KW	sarcoma; neurodegenerative; inflammatory; rheumatoid arthritis; asthma;
KW	multiple sclerosis; psoriasis; neuronal; Alzheimer's disease; dementia;
KW	depression; epilepsy; acquired immuno deficiency syndrome; allergy;
KW	AIDS; anaemia; atopic dermatitis; diabetes mellitus; dermatological;
KW	myocardial infarction; renal tubular acidosis; gonadal dysgenesis;
KW	dysplasia; cataract; cytostatic; neuroprotective; nootropic; anti-HIV;
KW	anticonvulsant; antiinflammatory; Cushings syndrome; cardiant;
KW	ophthalmological; transmembrane domain.
OS	Homo sapiens.
XX	
PN	WO200236741-A2.
XX	
PD	10-MAY-2002.
XX	
XX	29-OCT-2001; 2001WO-US46559.
XX	
PR	30-OCT-2000; 2000US-244688P.
PR	30-JUL-2001; 2001US-308706P.
XX	
XX	(BRIM) BRISTOL-MYERS SQUIBB CO.
XX	

CC differences in gene sequences among normal, carrier and affected
 CC individuals. Using diagnostic assays, cancer can be detected prior to
 CC the appearance of clinical symptoms and thereby progression of cancer can
 CC be prevented by aggressive treatment or preventive measures.

XX
 CC Sequence 308 AA;
 CC
 CC Query Match 50.3%; Score 1120.5; DB 21; Length 308;
 CC Best Local Similarity 53.7%; Pred. No. 9e-87;
 CC Matches 230; Conservative 29; Mismatches 41; Indels 85; Gaps 6;
 CC
 CC 41 LRRLKLESTFHRGHERFRITASACLDCEFLLAGAGAGAAAPGPHLPGRGSPVG 99
 CC 6 LKAEKFLASTFDRKNERFRIVSWKDLHCQFLVPOQG-----SPHSLPP----- 51
 CC
 CC 100 DPVRIHCNITESPAPPIWVSDDPNLAALVRLVDIKKGNFTLLQLHLKRIISDLCKL 159
 CC 52 -PLTLHCNITESPSPINFDSEDPNLASVLERLEDTKNNL----- 94
 CC
 CC 160 YNLPHQPDVEMLOPLPAEQCTQEDVSS--EDEDEMPEDTEDLDHYEMKEEPEAGKKS 217
 CC 95 -----NCTTEVTSSEEESEEAEDIEDLDHYEMKEEPEISGKKS 135
 CC
 CC 218 EDDGIGKENLAILEKIKKNOQDYLVNGAVSGSVQATDRLMKELDIYRSQSFKGNVAYE 277
 CC 136 EDEGIGKENLAILEKIKKNOQDYLVNGAVSGSVQATDRLMKELDIYRSQSFKGNVAYE 195
 CC
 CC 278 LVNDSLDVNWVKKLVQDQSDALHNDLQILKEKEGADFTLLNFSKDNFPDPFVRVVP 337
 CC 196 LINDSLVDHWVKKLVQDQSDALHNDLQILKEKEGADFTLLNFSKDNFPDPFVRVVP 255
 CC
 CC 338 VLSGGYVGGGAGICMELLTKOGWSSAYSIESVIMQISATLVKGRVQFGANKSQYSLTR 397
 CC 256 VLSGGYVGGGAGICMELLTKO-----NQYNLAR 283
 CC
 CC 398 AQQYSKSIQVQIHEKNGWTPPKEDG 422
 CC 284 AQQYSKSIQVQIHEKNGWTPPKEDG 308

RESULT 8
 ABB59516
 ID ABB59516 standard; Protein; 397 AA.
 XX
 AC ABB59516;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 5340.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-061415O.
 XX
 PA (PEKE) PE CORP NY.
 XX
 FI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL03619.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

Interactions -

PT
 XX Disclosure; SEQ ID NO 5340; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 397 AA;

Query Match 45.8%; Score 1020.5; DB 22; Length 397;
 Best Local Similarity 49.0%; Pred. No. 4.1e-78;
 Matches 202; Conservative 72; Mismatches 87; Indels 51; Gaps 4;
 CC
 CC 41 LRRLKLESTFHRGHERFRITASACLDCEFLLAGAGAGAAAPGPHLPGRGSPVG 100
 CC 7 LKAEKFLASTFDRKNERFRIVSWKDLHCQFLVPOQG-----DKNGK 45
 CC
 CC 101 PVRHCNITESPAPPIWVSDDPNLAALVRLVDIKKGNFTLLQLHLKRIISDLCKLY 160
 CC 46 RYDIHANITETYPSSPPVWFAESEEVTNAVQILSNTNGRDNHVINQVIGILLRELRLH 105
 CC
 CC 161 NLPQHPDVEMLDPLPAEQCTQEDVSEDE-----DEEMPEDTEDLDHYEM 206
 CC 106 NVLPDPDIDNALPLQTPPPSPALRCEQRPGGGAGGGGPHGNEETSDQBEIEDPIG 165
 CC
 CC 207 KEEPEAGKK-----SEDDGIGKENLAILEKIKKNOQDYLVNGAVSGSVQATD 254
 CC 166 ESEQSEGEDDLPLEMDVYRSTSKDDMEVEHLATLEKLRQSDYLVNGAVSGSVQATD 225
 CC
 CC 255 RLMKELDIYRSQSFKGNVAYVELVNDSDVWVKKLVQDQSDALHNDLQILKEKEGAD 314
 CC 226 RLMKELDIYRSQSFKGNVAYVELVNDSDVWVKKLVQDQSDALHNDLQILKEKEGAD 285
 CC
 CC 315 ILLNFSKDNFPDPFVRVVPVLSGGYVGGGAGICMELLTKOGWSSAYSIESVIMQIS 374
 CC 286 ILLNFSKDNFPDPFVRVVPVLSGGYVGGGAGICMELLTKOGWSSAYSIESVIMQIS 345
 CC
 CC 375 ATLVKGRVQFGANKSQYSLTRAQYSLVQIHEKNGWTPPKEDG 422
 CC 346 ATLVKGRVQFGANKSQYSLTRAQYSLVQIHEKNGWTPPKEDG 397

RESULT 9

AAE24524
 ID AAE24524 standard; Protein; 164 AA.

XX AAE24524;

AC AAE24524;

DT 04-OCT-2002 (first entry)

XX Human RAT1d6 ubiquitin conjugating enzyme (UBC) domain.

XX Human; ubiquitin conjugating enzyme; UBC; RAT1d6; immune disorder;
 KW regulated in activated T-lymphocyte 1d6; neuronal disorder; cancer;
 KW tumour; lymphoproliferative; cancer; adenocarcinoma; leukaemia; myeloma;
 KW sarcoma; neurodegenerative; inflammatory; rheumatoid arthritis; asthma;
 KW multiple sclerosis; psoriasis; neuroendocrine; Alzheimer's disease; dementia;
 KW depression; epilepsy; acquired immunodeficiency syndrome; allergy;
 KW AIDS; anaemia; atopic dermatitis; diabetes mellitus; dermatological;
 KW myocardial infarction; renal tubular acidosis; gonadal dysgenesis;
 KW dysplasia; cataract; cytostatic; neuroprotective; nontropic; anti-HIV;
 KW anticonvulsant; antiinflammatory; Cushing's syndrome; cardiac;
 KW ophthalmological; transmembrane domain.

OS Homo sapiens.
 XX WO200236741-A2.
 XX 10-MAY-2002.
 XX 29-OCT-2001; 2001WO-US46559.
 XX 30-OCT-2000; 2000US-244688P.
 XX 30-JUL-2001; 2001US-308706P.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Bowen MA, Wu Y, Yang W, Finger JN;
 XX WPI; 2002-479758/51.
 XX Novel ubiquitin conjugating enzyme polypeptide isolated from activated
 PT human T cell, for screening modulators useful for treating cancer,
 PT immune disorder, lymphoproliferative disorder, neurodegenerative
 PT disorder
 XX
 PS Disclosure; Page 167; 169pp; English.
 XX
 CC The invention relates to a novel ubiquitin conjugating enzyme (UBC)
 CC homologue, RATLid6 (regulated in activated T-lymphocytes lid6) and its
 CC corresponding nucleic acid. The invention also relates to methods for
 CC treating, diagnosing, preventing and screening for disorders related
 CC to the expression of RATLid6. UBC is useful for screening for candidate
 CC compounds capable of binding to and/or modulating its activity. UBC is
 CC useful for treating an immune or neuronal disorder in a mammal. The
 CC method is useful for treating a cancer or tumour. It is useful for
 CC suppressing the immune response in a subject requiring the suppression.
 CC It is also useful for treating lymphoproliferative disorder, cancer e.g.
 CC adenocarcinoma, leukaemia, myeloma, sarcoma, etc, neurodegenerative
 CC disorder, inflammatory disorders e.g. rheumatoid arthritis, asthma,
 CC multiple sclerosis, psoriasis, etc, neuronal disorders e.g. Alzheimer's
 CC disease, dementia, depression, epilepsy, etc, immune disorder or immune
 CC related disorders such as acquired immuno deficiency syndrome (AIDS),
 CC allergy, anaemia, atopic dermatitis, diabetes mellitus, myocardial
 CC infarction, etc, developmental disorders e.g. Cushing's syndrome, renal
 CC tubular acidosis, gonadal dysgenesis, dysplasia, cataract, etc. The
 CC present sequence is human RATLid6 UBC domain.
 XX
 XX Sequence 164 AA;
 SQ
 Query Match 37.4%; Score 834; DB 23; Length 164;
 Best Local Similarity 100.0%; Pred. No. 8.4e-63;
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 248 GSVQATDRLMKELRDYRSQSFKGGNYAVELVNDSLYDNVLLKVDQDSALHNDLIQILK 307
 DB 1 GSVQATDRLMKELRDYRSQSFKGGNYAVELVNDSLYDNVLLKVDQDSALHNDLIQILK 60
 QY 308 EREGADFIILNFSFKDNFFDPFVRVSPVLGGVGLGGATCMELLTKQGWSSAYSIE 367
 DB 61 EREGADFIILNFSFKDNFFDPFVRVSPVLGGVGLGGATCMELLTKQGWSSAYSIE 120
 QY 368 SVIMQISATLVKGRVQFGANKSQYSLTRAQSQYSLVQIHEK 411
 DB 121 SVIMQISATLVKGRVQFGANKSQYSLTRAQSQYSLVQIHEK 164
 RESULT 10
 ABB67021
 ID ABB67021 standard; Protein; 217 AA.
 XX
 AC ABB67021;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster polypeptide SEQ ID NO 27855.
 DE
 XX

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 XX (PEXE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX DR N-PSDB; ABL11124.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions
 XX
 PS Disclosure; SEQ ID NO 27855; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 217 AA;
 SQ
 Query Match 36.2%; Score 808; DB 22; Length 217;
 Best Local Similarity 72.4%; Pred. No. 2e-60;
 Matches 152; Conservative 31; Mismatches 23; Indels 4; Gaps 1;
 QY 217 SEDDGIGKENLAILEKIKKNQDYLNGAVSGSVQATDRLMKELRDYRSQSFKGGNYAV 276
 DB 8 SKKDMEVEHLATLEKLRQSRQDYLGKSGSVQATDRLMKELRDYRSQSFKGGNYAV 67
 QY 277 ELVNDSLYDNVLLKVDQDSALHNDLIQILKEGADFIILNFSFKDNFFDPFVRVVS 336
 DB 68 ELVNESIYEWNTLRKSVDPDSPDLHSDLOMLKEGKDSLILNLFKETYPPFEPFVRVH 127
 QY 337 PVLGGYVLGGATCMELLTKQGWSSAYSIESVIMQISATLVKGRVQFGANKS- ---Q 392
 DB 128 PIISGGYVLGGATCMELLTKQGWSSAYTVEAVIMQIATLVKGRVQFGATKALTQCG 187
 QY 393 YSLTRAQSQYSLVQIHEKNGWYTPPKEDG 422
 DB 188 YSLARAQSQFSLVQIHEKNGWYTPPKEDG 217
 RESULT 11
 ABB60719
 ID ABB60719 standard; Protein; 306 AA.
 XX
 AC ABB60719;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster polypeptide SEQ ID NO 8949.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW


```

Db      333 MLDQPLTGQNGTTEVTSEEEEE 357
        170 MLDQPLPAEQ 179.
        | | | | |
        183 MPDQPLPMGQ 192

RESULT 15
ABG15534
ID ABG15534 standard; Protein; 1909 AA.
XX AC ABG15534;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #15525.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS79721.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 45893; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1909 AA;

Query Match      8.4%; Score 187.5; DB 22; Length 1909;
Best Local Similarity 58.6%; Pred. No. 2.7e-06;
Matches 41; Conservative 7; Mismatches 21; Indels 1; Gaps 1;

OY 110 ESYPAVPPIWSESDPNLAVALPELVDIKKNTLLLOHLKRIISDLCKLYNLQHPDVE 169
    ||| : ||| | : ||| : || : ||| | | | | | | | | | | : ||| : |||
Db 124 ESYSSSPPIWFVSDLELNLTSLVECLEL-DRNNLLHQOLKWLKICGLRYLNPRLDVA 182

```

Search completed: July 29, 2003, 13:44:52
 Job time : 85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 13:42:28 ; Search time 42 Seconds
(without alignments)
966.266 Million cell updates/sec

Title: US-10-005-549-2

Perfect score: 2229

Sequence: 1 MQQPQPGQQQPGPGQQLGG.....KSLVQIHEKNGWYTPPKEDG 422

Scoring table: BLOSUM62

Gapop. 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: p1r1.*

2: p1r2.*

3: p1r3.*

4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	841	37.7	471	T21349	hypothetical prote
2	135.5	6.1	295	A41241	ubiquitin-conjugat
3	133	6.0	1300	T03166	probable immediate
4	132.5	5.9	1586	S39580	HRBM protein - hum
5	131	5.9	199	T08465	ubiquitin-protein
6	129.5	5.8	1572	S45251	SNFalpha protein
7	123.5	5.5	900	B70694	probable infb - My
8	122	5.5	924	S12746	potassium channel
9	121.5	5.5	147	F30082	ubiquitin conjugat
10	121.5	5.5	147	F30118	ubiquitin conjugat
11	121.5	5.5	147	F30137	ubiquitin conjugat
12	121.5	5.5	147	F30102	ubiquitin conjugat
13	121.5	5.5	147	H90116	ubiquitin conjugat
14	120.5	5.4	385	T20410	hypothetical prote
15	120.5	5.4	924	C86725	chromosome segrega
16	120.5	5.4	1613	S39059	protein BRG1 - hum
17	120	5.4	756	T30367	hypothetical prote
18	119	5.3	924	B41359	potassium channel
19	118.5	5.3	700	G70192	antigen, p83/100 h
20	118.5	5.3	1445	A59437	KIAA1204 protein f
21	117.5	5.3	590	A40437	glutamic acid-rich
22	117.5	5.3	1603	S23810	collagen alpha 1(X
23	117	5.2	5037	B35041	ryanodine receptor
24	116.5	5.2	1262	T33074	hypothetical prote
25	116.5	5.2	1558	B71603	RESA-H3 antigen pF
26	116	5.2	195	D96541	hypothetical prote
27	116	5.2	924	F87103	initiation factor
28	116	5.2	1647	S45252	SNF2beta protein -
29	115.5	5.2	5035	I46646	ryanodine receptor

30 115 5.2 871 1 TNBEA1
31 114.5 5.1 5032 1 A35041
32 114 5.1 156 2 S67248
33 114 5.1 519 2 S78089
34 114 5.1 1928 2 S46773
35 113.5 5.1 152 2 T27167
36 112.5 5.0 323 2 S20099
37 112.5 5.0 364 1 TVHUM1
38 112.5 5.0 1151 2 G96805
39 112 5.0 467 2 G84545
40 112 5.0 962 1 S57342
41 111.5 5.0 147 2 T23820
42 111.5 5.0 147 2 A48145
43 111.5 5.0 147 2 I39202
44 111.5 5.0 751 2 T48719
45 111.5 5.0 1063 2 T03743

ALIGNMENTS

RESULT 1

T21349

hypothetical protein F25H2.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T21349

R:Wilkinson, J.

submitted to the EMBL Data Library, September 1996

A:Reference number: Z19411

A:Accession: T21349

A:Status: preliminary; translated from GB/EMBL/DBJ.

A:Molecule type: DNA

A:Residues: 1-471 <WIL>

A:Cross-references: EMBL:Z79754; PIDN:CA802096.1; GSPDB:GN00019; CESP:F25H2.8

A:Experimental source: clone F25H2

C:Genetics:

A:Gene: CESP:F25H2.8

A:Map position: 1

A:Introns: 26/3; 138/3; 302/3; 353/3; 379/2

C:Superfamily: Caenorhabditis elegans hypothetical protein F25H2.8

Query Match 37.7%; Score 841; DB 2; Length 471;
Best Local Similarity 44.9%; Pred. No. 6.9e-46;
Matches 180; Conservative 69; Mismatches 100; Indels 52; Gaps 10;

Qy 40 CLRR---ELKLESIFRHERFERIASACDELSCFELLACAGAGAGAGAGPGLPRGS 96

Db 3 CLRLKEDIQVLEKLFKPNHNRFOILSASVDELSMKFINAENKG----- 46

Qy 97 VPGDPVRIHCNITESYPVPPVWSVESDD-PNLAALVRLVDIKKNTLLQLHLKRIIS 155

Db 47 -----IIVTANIQENYPRQPIWFSSEDDVPVIGMSLQRLTETEE-STNLHQVRLVSD 100

Qy 156 LCKLYNL-----POHPDV-----EMLDQPLPAEQCTQEDVSSDEDEEMP 195

Db 101 LCSFYNLQMPCELPQIAPPVRDDIDEGRGSDISDTTSEPIDDDMAGDGEVDDDEEDD 160

Qy 196 EDYE-DLDHVMKEEPEAEKSKSDDGIGENLAILEKIKKNORODYLNAGVSGVQATD 254

Db 161 EDADGDIEIVEMAEEDPT----SQHDVGVSKEGDMLDKVSKNRQQLDGVKGSGITATD 217

Qy 255 RLKRELDIRVSQFGKGNAYVEL-VNDSLYDNVKKLKYDDDSALHDLQILKEKGD 313

Db 218 RLKKEIDIRHSEHFKNQITFELEKENLYQWIKLKHVDEDSPLFEDMKKLLKHQND 277

Qy 314 FILNFSKDNFPDPFVRVSVLGGVGLGGATCMELLTKQGWSSAYSIESVIMQI 373

Db 278 HLLFSFTFNEKFPDPPFVRVAVPHINQGVGLGGATCMELLTKQGWSSAYSIESVIMQI 337

Qy 374 SATLVKQKARVQGA-NKQSYSLTRAQQSYKSLVQIHEKNG 413

Db 338 AATLVKGRARISFDAKHTSTYSMARAAQSQFSLQQLIHAQSG 378

97K alpha trans-in
ryanodine receptor
hypothetical prote
G-protein signalin
myosin heavy chain
hypothetical prote
transforming prote
transforming prote
hypothetical prote
bHLH transcription
endopeptidase La h
hypothetical prote
ubiquitin-conjugat
ubiquitin-conjugat
hypothetical prote
bifocal protein -

	Query Match	6.1%;	Score 135.5;	DB 2;	Length 295;
	Best Local Similarity	21.4%;	Pred. No. 0.19;		
	Matches	42;	Conservative	33;	Mismatches 62;
				Indels	59; Gaps 7;
QY	273 NYAVELNDS-LYDNNVKLLKLVDDSDALHNDLQILKEGADFIILLNFSPFDKNPFDPDPF	331	:	:	:
Db	28 SFHLEEDDSNIFTWNTGVVLNEDSIY-----GGFFKAQRFPDEDFPFPQPQ	76	:	:	:
QY	332 VRVSPVLSGGYVLGGGAICMELLTKQG-----WSAYSIESVIMQISATL----	377	:	:	:
Db	77 FR-FTPAYIHPNVVRDGRLCISILHQSGDPWTDEPTDAETWSPVQTVESVLISIVSLLED	135	:	:	:
QY	378 -----VKGKARVQFGANKSQYS-----LTRAQQSYKSULVQIHE----	410	:	:	:
Db	136 NINSPANVDRAVDYTRKNPEQYQKRVKMEVERSKODIPKGFIMPTSESAYISQSXLDEPES	195	:	:	:
QY	411 -----KNGWYTTPPKED	421	:	:	:

A;Accession: S67870
A;Molecule type: DNA
A;Residues: 1-295 <BLO>
A;Cross-references: EMBL:Z74350; NID:gi431496; PID:gi431497; MIPS:YDR054C
C;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:CDG34; UBC3
A;Cross-references: SGD:S0002461; MIPS:YDR054C
A;Map position: 4R
C;Keywords: nucleus

QY	273	NYAVELVND\$-LYDNWNKVLKVOD\$SALHNDLQILKEGAD\$FILLNF\$FKDNFP\$FDP\$P\$ 331
Db	28	\$F\$HLEDD\$NIFT\$WIG\$VMV\$NED\$SI\$V\$-----GG\$FKAQMRP\$FDEP\$F\$P\$Q 76
QY	332	VRV\$P\$VLSGGYVLGGCAICHEL\$TKQ\$-----W\$SAYSIESVIMQISAT\$L\$--- 377
Db	77	FR-\$F\$T\$P\$AI\$Y\$H\$N\$V\$Y\$R\$D\$R\$G\$L\$C\$IS\$I\$L\$H\$Q\$D\$P\$M\$T\$D\$E\$F\$A\$T\$W\$S\$P\$V\$O\$T\$V\$S\$V\$S\$LL\$E\$D\$P 135
QY	378	-----VKGK\$K\$V\$Q\$F\$G\$ANK\$Q\$S\$Y\$-----L\$T\$R\$A\$Q\$Q\$Y\$K\$S\$V\$O\$I\$H\$E\$--- 410
Db	136	NIN\$P\$AN\$V\$D\$R\$A\$V\$Y\$R\$K\$N\$P\$E\$Q\$IK\$Q\$R\$K\$M\$E\$V\$E\$R\$S\$K\$O\$D\$IP\$K\$G\$F\$IM\$P\$T\$E\$S\$A\$Y\$IS\$Q\$S\$K\$L\$D\$E\$P\$S 195
QY	411	-----KNGW\$T\$P\$P\$K\$D 421

Query Match	5.93;	Score 132.5;	DB 2;	Length 1586;
Best Local Similarity	22.0%;	Pred. No. 2.8;		
Matches 68;	Conservative	39;	Mismatches 109;	Indels 93; Gaps 14;
QY	3	QPQPGGQQP-----GPGQQLGQGA-----	APGAGGGGPGGPGPCLRELKLL	48
DB	241	QPQTQQQQPALVNNRPSGPGBELSGSPQKLPVP-----	APGGRSPAP-----	287

QY 49 ESIFRHRHERFIASACLDLSCEFLLAGAGAGAGAAPGPHLPVRGPGDPVRIHCNI 108
 Db 288 -----PAAQPPAAAVPGSVPO--PAPQPSVP-LQL 317
 QY 109 TESYPVPPIVSVESDDNLAVALERLVLDIKKNTLLQHLKRIISDLCKLYNLPHQ--P 166
 Db 318 QOKOSRISPIOKPQGLDP-----VEILQREYRLQ--ARIAHRIQELNLPGLSLPP 366
 QY 167 D-----VEMLDQPLP--AEQCTQEDVSSEDEEMPEDETLQHYEMKEEPAEGKKSE 218
 Db 367 DLRTKATVELKALRLNFORQLREEVVACMRDITLETALNSKAYKRSKQTLREARMTE 426
 QY 219 DDGIGKENLAILEKIKKNQODYLNAGVSGSVQATDRMLKELRDYRSQSFKGGNYAVEL 278
 Db 427 K--LEKQOKIEQERKRQKHQEYLSILOHA-----KDFKEYHRSVAGK-----IQK 471
 QY 279 VNDSLYDWN 287
 Db 472 LSKAVATWH 480

RESULT 5
 T08465
 ubiquitin-protein ligase (EC 6.3.2.19) - fruit fly (Drosophila melanogaster)
 N:Alternate names: ubiquitin conjugating enzyme
 C:Species: Drosophila melanogaster
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 03-Jun-2002
 C:Accession: T08465
 R:Kirby, R.J.
 submitted to the EMBL Data Library, June 1996
 A:Reference number: 216421
 A:Accession: T08465
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-199 <KIR>
 A:Cross-references: EMBL:X92838; NID:g1359613; PID:g1054722
 A:Experimental source: strain Canton-S
 C:Genetics:
 A:Gene: UbcD4
 A:Cross-references: FlyBase:FBgn0015321
 C:Superfamily: yeast ubiquitin-protein ligase UBC1
 C:Keywords: cell cycle control; DNA repair; ligase; protein degradation

Query Match 5.9%; Score 131; DB 2; Length 199;
 Best Local Similarity 24.7%; Pred. No. 0.21;
 Matches 38; Conservative 32; Mismatches 56; Indels 28; Gaps 7;

QY 252 ATDRMLKELRDYRSQSFKGGNYAVELVNDSLYDWNVKKLLKVDODSALHNDLQILKEK-- 309
 Db 5 AVSRIKREFKVMRSEETVQCSIKIELVNDSD--W-----TELGEIAGPPDTPY 51
 QY 310 EGADFILLNFSKDNFPDPFVRVSPVLGGVGLGGATCMELLTKQGWSSAYSIESV 369
 Db 52 EGGKFV-LEIKVPEYFPNPKARFTRIHNPNISSVTGAICLDIL-KDNWAAAMLTRTV 109
 QY 370 IMQISATLVKKG-----ARVOFGANKSOYSL 395
 Db 110 LLSLQALLAAEPPDQDQAVVAYQF---KDKYDL 140

RESULT 6
 S45251
 SNF2alpha protein - human
 C:Species: Homo sapiens (man)
 C>Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
 C:Accession: S45251
 R:Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.
 Nucleic Acids Res. 22, 1815-1820, 1994
 A:Title: Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brahm
 A:Reference number: S45251; MUID:94268902; PMID:8208605
 A:Accession: S45251
 A:Status: preliminary
 A:Molecule type: mRNA

A:Residues: 1-1572 <CHI>
 A:Cross-references: GB:D26155; NID:g505086; PIDN:BAA05142.1; PID:d1005684; PID:g9876
 C:Superfamily: human SNF2alpha protein; bromodomain homology
 F:1409-1464/Domain: bromodomain homology <BRO>

Query Match 5.8%; Score 129.5; DB 2; Length 1572;
 Best Local Similarity 21.7%; Pred. No. 4.3;
 Matches 67; Conservative 40; Mismatches 109; Indels 93; Gaps 14;

QY 3 QPQGGQGOQP-----GPGQOLGGOGA-----APGAGGGGGGPGPGPCLRLRELKLL 48
 Db 245 QPQTOQQOQPALVNNRPSGPGPELSPSTPQKLPVP-----APGGRSPAP----- 291
 QY 49 ESIFRHRHERFIASACLDLSCEFLLAGAGAGAGAAPGPHLPVRGPGDPVRIHCNI 108
 Db 292 -----PAAQPPAAAVPGSVPO--PAPQPSVP-LQL 321
 QY 109 TESYPVPPIVSVESDDNLAVALERLVLDIKKNTLLQHLKRIISDLCKLYNLPHQ--P 166
 Db 322 QOKOSRISPIOKPQGLDP-----VEILQREYRLQ--ARIAHRIQELNLPGLSLPP 370
 QY 167 D-----VEMLDQPLP--AEQCTQEDVSSEDEEMPEDETLQHYEMKEEPAEGKKSE 218
 Db 371 DLRTKATVELKALRLNFORQLREEVVACMRDITLETALNSKAYKRSKQTLREARMTE 430
 QY 219 DDGIGKENLAILEKIKKNQODYLNAGVSGSVQATDRMLKELRDYRSQSFKGGNYAVEL 278
 Db 431 K--LEKQOKIEQERKRQKHQEYLSILOHA-----KDFKEYHRSVAGK-----IQK 475
 QY 279 VNDSLYDWN 287
 Db 476 LSKAVATWH 484

RESULT 7
 B70694
 Probable infB - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 02-Feb-2001
 C:Accession: B70694
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gene
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: B70694
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-900 <COL>
 A:Cross-references: GB:281331; GB:ALI23456; NID:g3261650; PIDN:CAB03670.1; PID:g1648E
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: infB
 C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu
 C:Keywords: GTP binding; nucleotide binding; P-loop
 F:399-512/Domain: translation elongation factor Tu homology <ETU>
 F:405-412/Region: nucleotide-binding motif A (P-loop)
 F:509-512/Region: GTP-binding NKXD motif
 F:545-547/Region: GTP-binding SAK/L motif
 F:411,412,432,509,510,512,545/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser

Query Match 5.5%; Score 123.5; DB 2; Length 900;
 Best Local Similarity 20.5%; Pred. No. 4.9;
 Matches 92; Conservative 63; Mismatches 128; Indels 165; Gaps 21;

QY 4 PQQGGQ-QPQPGQ-----OLGGQGAAPGAG--GGPGGGPGPGCLRLRELKLLSIF 52
 Db 211 PRPGRGPAPGACGRSDAGGNGYRGVGAAPGTGFRGPRGGGGGPRGQR----- 260
 QY 53 HRGHERFIASACLDLSCEFLLAGAGAGAGAAPGPHLPVRG----- 95

Db 261 -----GGAAGAFRPG-GAPRRGRKSRKQKROEYDSWQAP 294
Qy 96 -----SVP-----GDPVRIHCNITESYPAPPIWSVESDDPNLAALVRLVDIKKNTLLQ 147
Db 295 VVGVRPHNGETIRLARGASLSDF-----DKIDANPAALVQALFNIGE-----MVT 343
Qy 148 HLKRIISDLCKLYNLPQHPVMDQLPABOCTQEDVSEDEDEEMPEDEDLHYEMK 207
Db 344 ATQSVGDETLELGSSEMYNNVQV-----VSPDEDELELLE-SFDLSY----- 384
Qy 208 EEPAPGKSKSDDGIGKENLAI-----LEKIKK-NQODYLYNG----- 244
Db 385 -----GDEG-GEDLQVRPPVTVMGVHDHGTGKRLDTTRKANVREAEAGGITQH 434
Qy 245 -----AVSGSVQATDRLM-----KELRDYRSQSFSGYNAVVELNDSDLYDMNVKLLV 293
Db 435 IGAYQVAVDLDSGQRLTFTDPGHEAFTAMRAGAKATDIAI-----LVVA 481
Qy 294 DODSALHNDLQILKEGADP-ILLNFSKDNFFDPFVVRVSVLGGVY-----LGGGA 349
Db 482 ADDGVMPQTVRAINHAQAADVPYVAVNKKIDEGADPAKIR--GOLTEYGLVPEEFGDT 539
Qy 350 ICMELLTKQGWSSAYSIESVIMQISATL 377
Db 540 MFVDSAKQGTNIEALEEAVLTTADAL 567

RESULT 8

S12746
potassium channel protein shab11 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Aug-1999
C:Accession: S12746; PMID:90243668; PMID:2336395
R:Butler, A.; Wei, A.; Salkoff, L.
Nucleic Acids Res. 18, 2173-2174, 1990
A:Title: Shal, Shab, and Shaw: three genes encoding potassium channels in Drosophila.
A:Reference number: S12746; PMID:90243668; PMID:2336395
A:Accession: S12746

A:Molecule type: mRNA
A:Residues: 1-924 <BUT>
A:Cross-references: EMBL:M32659
R:Butler, A.; Wei, A.; Salkoff, L.
submitted to the EMBL Data Library, March 1990
A:Description: Shal, Shab, and Shaw: Three genes encoding potassium channels in Drosophila.
A:Reference number: S15058
A:Accession: S15058

A:Molecule type: mRNA
A:Residues: 1-625, 'YG', 628-924 <BUT>
A:Cross-references: EMBL:M32659; NID:gl58458; PIDN:AAA28896.1; PID:gl58459
C:Genetics:

A:Gene: shab11
A:Cross-references: FlyBase:FBgn0003383
C:Superfamily: potassium channel protein drk1
C:Keywords: ion channel; potassium channel; transmembrane protein

Query Match 5.5%; Score 122; DB 2; Length 924;
Best Local Similarity 23.9%; Pred. No. 6.3;
Matches 64; Conservative 27; Mismatches 101; Indels 76; Gaps 12;

Qy 13 GPGQOLGGGAAPGAGGPGGPGP-CLRELKLLSIFRHERFRIASACLD-ELS 70
Db 163 GAGASVTGGSGAGTGTGTAGSGSGGGAAGKEVR-----YAPFPVASPTHSIPTT 213
Qy 71 CEFLLAGAGAGAGAAPGPHLPGRGVP-----GDPVRIHCNITESYPAVPPINSVE 122
Db 214 SQQIVGGVGVGGASSQSI--SGVPTHSSQNTGTALQTHSRMSISIPPPFFMTAQ 271
Qy 123 SDDPN-----LAALVRLVDIKKNTLLQHLKRIISDLCKLYN----- 162
Db 272 SKAVNSRVSINVGVRVHVRVLRPLRLPLRLGLRGCTTHEAIV-ELCDDYSLADNEYF 330
Qy 163 -PQHP-----DVEMLDQ-----PLPAEQCTQEDVSEDEBD- 191

Db 331 FDRHPKFSFSSILNFYRTGKLHIYDEMCVLAFGDDLEWGVDELYLESCQHKYHQKENV 390
Qy 192 -EEMPEDETLDHYEMKEEPAEGKKE 218
Db 391 HEEMKEAESL--RQRDEEEFGGKFE 416

RESULT 9

F90082
ubiquitin conjugating enzyme [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Jun-2001
C:Accession: F90082
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.;
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; MUID:11323671; PMID:11323671
A:Accession: F90082
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-147 <DOU>
A:Cross-references: GB:AF165818; NID:g6690600; PIDN:AAF24208.1; GSPDB:GN00150
C:Genetics:

A:Gene: ubc4
A:Map position: 1
A:Genome: nucleomorph
C:Superfamily: human ubiquitin-protein ligase E2
C:Keywords: nucleomorph
Query Match 5.5%; Score 121.5; DB 2; Length 147;
Best Local Similarity 25.3%; Pred. No. 0.56; 59; Indels 27; Gaps 7;
Matches 39; Conservative 29; Mismatches 59;

Qy 252 ATDLRMLKELRDYRSQSFSGYNAVVELNDSDLYDMNVKLLKVPQDSALHNDLQILKEG 311
Db 2 ATRIKQEKHDLKDDP--SNCAGSPDENLFRWFIASIG-FSDSPY-----A 46
Qy 312 ADFILNFSKDNFFDPFVVRVSVLGGVYLVGGGAICMELLTKQGWSSAYSIESVIM 371
Db 47 GGVFFLSILFPDYPFPKPKVQFTTKIYHPN-INSNGSICLDILRDQ-WSPALTVSKVLL 104
Qy 372 QISA-----TLVKGKARVQFGANKSQYSLT 396
Db 105 SISLLTDPNDPDDPLVPEIARV-FKENSRYEAT 137

RESULT 10

F90118
ubiquitin conjugating enzyme [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Jun-2001
C:Accession: F90118
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.;
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; MUID:11323671; PMID:11323671
A:Accession: F90118
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-147 <DOU>
A:Cross-references: GB:AF083031; NID:g6690139; PIDN:AAF24004.1; GSPDB:GN00152
C:Genetics:

A:Gene: ubc4
A:Map position: 3
A:Genome: nucleomorph
C:Superfamily: human ubiquitin-protein ligase E2
C:Keywords: nucleomorph
Query Match 5.5%; Score 121.5; DB 2; Length 147;
Best Local Similarity 25.3%; Pred. No. 0.56; 59; Indels 27; Gaps 7;
Matches 39; Conservative 29; Mismatches 59;

[illegible]

search completed:
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 13:35:48 ; Search time 23 Seconds
(without alignments)

862.838 Million cell updates/sec

Title: US-10-005-549-2

Perfect score: 2229

Sequence: 1 MQQPQPGQQPGGQOLGG.....KSLVQIHKNGWTPPKEDG 422

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135.5	6.1	295	1	UBC3_YEAST
2	132.5	5.9	1586	1	SN22_HUMAN
3	129.5	5.8	207	1	UB6B_MOUSE
4	125.5	5.6	1256	1	SN3_RAT
5	123.5	5.5	900	1	IF2_MYCTU
6	123	5.5	199	1	UBCB_DROME
7	122.5	5.5	177	1	UBCB_SPI50
8	122.5	5.5	1048	1	AGOL_ARATH
9	121.5	5.5	193	1	UBC6_MOUSE
10	120	5.4	756	1	EFER_HUMAN
11	120	5.4	917	1	GCFC_HUMAN
12	119.5	5.4	193	1	UBC6_HUMAN
13	118.5	5.3	1210	1	BAT8_HUMAN
14	118	5.3	985	1	CIKB_DROME
15	117.5	5.3	179	1	UBCB_XENLA
16	117.5	5.3	1394	1	CNG4_BOVIN
17	117.5	5.3	1603	1	CAIF_HUMAN
18	117	5.2	5037	1	RYR1_RABIT
19	116.5	5.2	1263	1	BAT8_MOUSE
20	116	5.2	924	1	IF2_MYCLE
21	116	5.2	1647	1	SN24_HUMAN
22	115.5	5.2	671	1	HAP1_HUMAN
23	115.5	5.2	5035	1	RYR1_PIG
24	115	5.2	871	1	UL47_HSVB
25	114.5	5.1	5038	1	RYR1_HUMAN
26	114	5.1	156	1	UBCA_YEAST
27	114	5.1	519	1	RG53_HUMAN
28	114	5.1	1928	1	MTSL_YEAST
29	113	5.1	179	1	UBC2_CANAL
30	112.5	5.0	152	1	UBC2_MEDSA
31	112.5	5.0	323	1	JUND_CHICK
32	112.5	5.0	364	1	MYCL_HUMAN
33	112.5	5.0	739	1	DAXX_MOUSE

34	111.5	5.0	147	1	UB5A_HUMAN	P51668	homo sapien
35	111.5	5.0	147	1	UBC2_CAEEL	P35129	caenorhabdi
36	111.5	5.0	350	1	USX1_CHICK	Q91A12	gallus gall
37	111.5	5.0	738	1	S521_RAT	Q9QY02	rattus norv
38	111.5	5.0	740	1	DAXX_HUMAN	Q9UER7	homo sapien
39	111	5.0	150	1	UBC3_ARATH	P42746	arabidopsis
40	110.5	5.0	148	1	UBC5_YEAST	P15732	saccharomyc
41	110.5	5.0	148	1	UBCE_ARATH	P35134	arabidopsis
42	110.5	5.0	448	1	YAP1_CHICK	P46936	gallus gall
43	110.5	5.0	704	1	RP3A_BOVIN	Q06846	bos taurus
44	110	4.9	628	1	HAP1_MOUSE	Q35668	mus musculu
45	109.5	4.9	147	1	UBC1_DROME	P25867	drosophila

ALIGNMENTS

RESULT 1

UBC3_YEAST	STANDARD;	PRT;	295 AA.
ID	UBC3_YEAST		
AC	P14682;		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DT	01-APR-1990 (Rel. 14, Last annotation update)		
DE	UBiquitin-conjugating enzyme E2-34 kDa (EC 6.3.2.19)		
DE	(UBiquitin-protein ligase) (UBiquitin carrier protein) (Cell division control protein 34).		
GN	UBC3 OR CDC34 OR DNA6 OR YDR054C OR YP9609.08C OR D4211.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID:4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-88321694; PubMed-2842867;		
RA	Goelbl M.G., Vochem J., Jentsch S., McGrath J.P., Varshavsky A., Byers B.;		
RA	"The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme.";		
RL	Science 241:1331-1335(1988).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-96381250; PubMed-8789263;		
RA	Brandt P., Ramlow S., Otto B., Bloeker H.;		
RT	"Nucleotide sequence analysis of a 32,500 bp region of the right arm of Saccharomyces cerevisiae chromosome IV.";		
RL	Yeast 12:85-90(1996).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-S288c / AB972;		
RA	Hunt S., Bowman S., Harris D., Barrell B.G., Rajandream M.A.;		
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER PROTEINS. CAPABLE, IN VITRO, TO UBIQUITINATE HISTONE H2A.		
CC	-1- FUNCTION: MEDIATES THE INITIATION OF DNA REPLICATION (TRANSITION OF G1 TO S PHASE IN CELL CYCLE).		
CC	-1- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine - AMP + diphosphate + protein N-ubiquityllysine.		
CC	-1- PATHWAY: Ubiquitin conjugation; second step.		
CC	-1- SUBCELLULAR LOCATION: Nuclear.		
CC	-1- DOMAIN: THE ACIDIC C-TERMINAL EXTENSION IS ESSENTIAL FOR THE CELL CYCLE FUNCTION.		
CC	-1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-thiolester formation.		
CC	-1- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.		

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch).

```

CC EMBL; M21877; AAA35188.1; -
DR EMBL; X84162; CAA58970.1; -
DR EMBL; X74350; CAA98872.1; -
DR EMBL; Z49209; CAA89083.1; -
DR PIR; A41241; A41241.
DR HSP; Q02159; 20C3.
DR SGD; S0002461; CDC34.
DR GO; GO:0000086; P:G2/M transition of mitotic cell cycle; IPI.
DR InterPro; IPR000608; UBQ_conjugat.
DR Pfam; PF00179; UQ_con; 1.
DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
DR PROSITE; PS00127; UBIQUITIN_CONJUGAT_2; 1.
KW ubl conjugation pathway; Ligase; DNA replication; Nuclear protein;
KW Cell cycle; Cell division; Multigene family.
FT BINDING 95 95 UBIQUITIN (BY SIMILARITY).
FT DOMAIN 191 289 ASP/GLU-RICH (ACIDIC).
FT SEQUENCE 295 AA; 34064 MW; 1CE3E0C3AB1436DC CRC64;
SQ
Query Match 6.1%; Score 135.5; DB 1; Length 295;
Best Local Similarity 21.4%; Pred. No. 0.11;
Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;
QY 273 NYAVELVNDVLDYNNVKKLVQDSALHNDLQLKEGADFILLNFSFKDNFPDPFP 331
Db :: :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
28 SFHLEDDSNFTWNGVWLNDSIYH-----GGFFKAQMRFPEDFPSPQ 76
QY 332 VRVSPVLGGVYLGGAICMELITKQG-----WSAYSIESVIMQISATL--- 377
Db | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
77 FR-FTPAIYHPNVYDGRGLCSILHSGDPMTPDEPDATWSPVQVESVLISVSLLEDP 135
QY 378 -----VKGARVQFCANKSOKS-----LTRAQSQYSKLVQIHE--- 410
Db | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
136 NINSPANVDAVDYKRNPEYQKRVKMEVERSKODIPKGFIMPTSESAYISOSKLDEPES 195
QY 411 -----KNGWYTPPKED 421
Db || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
196 NKDMADNEWYDSLDLD 211
RESULT 2
ID SN22_HUMAN STANDARD; PRT; 1586 AA.
AC P51531;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Possible global transcription activator SNF2L2 (SNF2-alpha).
GN SMARCA2 OR SNF2L2 OR BRM OR SNF2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94038910; PubMed=8223438;
RA Muchardt C., Ivan M.;
RT "A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and
RT Drosophila brm genes potentiates transcriptional activation by the
RT glucocorticoid receptor.";
RL EMBO J. 12:4279-4290(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94268902; PubMed=8208605;
RA Chiba H., Muramatsu M., Nomoto A., Kato H.;
RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and
RT Drosophila brama are transcriptional coactivators cooperating with
RT the estrogen receptor and the retinoic acid receptor.";
RL Nucleic Acids Res. 22:1815-1820(1994).

```

```

CC -! FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR
CC HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
CC -! SUBCELLULAR LOCATION: Nuclear.
CC -! ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name-Long;
CC IsoId-P51531-1; Sequence-Displayed;
CC Name-Short;
CC IsoId-P51531-2; Sequence-VSP_000577;
CC -! SIMILARITY: Contains 1 bromodomain.
CC -! SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X72889; CAA51407.1; -
CC EMBL; D26155; BAA05142.1; -
CC PIR; S39580; S39580.
CC Genew; HGNC:11098; SMARCA2.
CC MIN; 600014; -
CC GO; GO:0005654; C:nucleoplasm; TAS.
CC GO; GO:0003713; F:transcription co-activator activity; TAS.
CC GO; GO:0006357; P:regulation of transcription from Pol II pro. . . ; TAS.
CC InterPro; IPR005576; BRK.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR006562; HSA.
CC InterPro; IPR000330; SNF2_N.
CC Pfam; PF00439; bromodomain; 1.
CC Pfam; PF00271; Helicase_C; 1.
CC Pfam; PF00176; SNF2_N; 1.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00592; BRK; 1.
CC SMART; SM00297; BROMO; 1.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELICC; 1.
CC SMART; SM00573; HSA; 1.
CC PROSITE; PS00633; BROMODOMAIN_1; 1.
CC PROSITE; PS0014; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW ATP-binding; Helicase; Alternative splicing.
FT DOMAIN 216 238 POLY-GLN.
FT DOMAIN 241 249 POLY-GLN.
FT DOMAIN 555 558 POLY-ARG.
FT DOMAIN 639 646 POLY-GLU.
FT NP_BIND 745 752 ATP (POTENTIAL).
FT SITE 847 850 DEGH BOX.
FT DOMAIN 1293 1297 POLY-GLU.
FT DOMAIN 1415 1485 BROMODOMAIN.
FT DOMAIN 1514 1525 POLY-GLU.
FT VARSPLIC 1397 1414 Missing (in isoform Short).
FT FTID=VSP_000577.
FT CONFLICT 239 239 P -> PQQP (IN REF. 2).
FT CONFLICT 390 390 Q -> E (IN REF. 2).
FT CONFLICT 509 509 G -> S (IN REF. 2).
FT CONFLICT 707 707 W -> R (IN REF. 2).
FT CONFLICT 1135 1135 D -> H (IN REF. 2).
FT CONFLICT 1390 1390 C -> V (IN REF. 2).
SQ SEQUENCE 1586 AA; 180762 MW; FA537F2A392807A CRC64;
Query Match 5.9%; Score 132.5; DB 1; Length 1586;
Best Local Similarity 22.0%; Pred. No. 1.6;
Matches 68; Conservative 39; Mismatches 109; Indels 93; Gaps 14;
QY 3 QPQPGQQQP-----GFGQQLGGQGA-----APGAGGGGGGGPGPCLRLKLL 48
Db ||| ||||| ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
241 QPQTQQQQQPALVNVNRPSPGPELSPGPKLPVP-----APGGRPSAP----- 287

```

QY 49 ESIFRHERPRIASACDELSCFELLAGAGGAGAGPGLPRLPPRSVPGDPVRIHCNI 108
 Db 288 -----PAAQPPAAAVPGSPVQ--PAPGQSPV-LQL 317
 QY 109 TESTPAYPPINWSDPDLNLAVERLVKIKKNTLLIQLHLKRIISDLCKLYNLPQH--P 166
 Db 318 QOKQSRISPIOKPQGLDP-----VEILQEREYRLQ--ARIAHRIQELNLPGLSPP 366
 QY 167 D-----VEMLDQPLP--AEQCTQEDVSSDEDEMPEDTDLHYEMKEEPAEGKKSE 218
 Db 367 DLRTKATVELKALLLNFQRLQEVVACMRDPTLETALNSKAYKRSKQTLREARWTE 426
 QY 219 DDGIGKENLALTEIKKRNORDYLVNGAVSSVQATDRMLKELRDYISQSFKGNYAVEL 278
 Db 427 K--LEKQOKEQEERKRQKQEQYLSILOH-----KDFEYHRSVAGK-----IQK 471
 QY 279 VNSLDYDWN 287
 Db 472 LSKAVATWH 480

RESULT 3
 UB6B_MOUSE STANDARD; PRT; 207 AA.
 AC P52483; O09180;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin-conjugating enzyme E2-23 kDa (EC 6.3.2.19)
 DE (Ubiquitin-protein ligase) (Ubiquitin carrier protein).
 GN UBCE4 OR UBCE2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss;
 RX MEDLINE=99087339; PubMed=9872334;
 RA Pestov D.G., Grzeszkiewicz T.M., Lau L.F.;
 RT "Isolation of growth suppressors from a cDNA expression library."
 RL Oncogene 17:3187-3197(1998).
 CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
 CC PROTEINS.
 CC -1- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine - AMP +
 CC diphosphate + protein N-ubiquityllysine.
 CC -1- PATHWAY: Ubiquitin conjugation; second step.
 CC -1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-
 CC thiolester formation (By similarity).
 CC -1- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X92664; CAA63352.1; --
 CC EMBL; AF003346; AAB60948.1; --
 CC HSSP; P15731; IQCQ.
 CC MGD; MGI:107412; Ubce4.
 CC InterPro; IPR000608; UbQ_conjugat.

DR Pfam; PF00179; UbQ_con; 1.
 DR ProDom; PD000461; UbQ_conjugat; 1.
 DR SMART; SMO0212; UBCC; 1.
 DR PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
 DR PROSITE; PS00127; UBIQUITIN_CONJUGAT_2; 1.
 KW Ubl conjugation pathway; Ligase; Multigene family.
 FT BINDING 145 145 UBIQUITIN (By similarity).
 FT CONFLICT 31 31 K -> E (IN REF. 2).
 SQ SEQUENCE 207 AA; 22911 MW; 86BCB59C8A72763F CRC64;
 Query Match 5.8%; Score 129.5; DB 1; Length 207;
 Best Local Similarity 24.2%; Pred.No. 0.17; 93; Indels 39; Gaps 10;
 Matches 53; Conservative 34; Mismatches 93; Indels 39; Gaps 10;
 QY 187 SEDEMPEDTDLHYEMKEEPAEGKKSEDDGIGKENLALTEIKKRNORDYLVNGAV 246
 Db 8 SDDESPSSGSSDAD-----QRDPAAPKPEEQ-----EERKPSATQOKKNTKL 51
 QY 247 GGSVQA-----TDRLMKELRDYRSQSFKGGNVAVELVNSLDYDWNVKKLVDDQSALHN 301
 Db 52 SSKTTAKLSTSAKRIQKELAEITLDP-----PNC-SAGPKGDNIYEWKSTIL--GPGSVY- 105
 QY 302 DLQILKEGEGADFIILNFSKDNFPDPFVRVSPVLSSGVVGLGGATCMELLTKQGS 361
 Db 106 -----EGGVF-FLDITFSSDYPPKPKVTFRTIYHCN-INSQGVICLDIL-KDNWS 154
 QY 362 SAYSIESVIMQISATLVK-GKARVQFGANKSQYSILTRAQ 399
 Db 155 PALTISKVLLSICSLTDCNPADPLVGSIAQTQVLTNRAE 193

RESULT 4
 SA3_RAT STANDARD; PRT; 1256 AA.
 ID SA3_RAT
 AC Q09M76;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cohesin subunit SA-3 (Stromal antigen 3) (SCC3 homolog 3).
 GN STAG3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar;
 RX MEDLINE=21482572; PubMed=11599053;
 RA Bayes M., Prieto I., Noguchi J., Barbero J.L., Perez Jurado L.A.;
 RT "Evaluation of the Stag3 gene and the synaptonemal complex in a rat
 RL model.(as/as) for male infertility."
 RL Mol. Reprod. Dev. 60:414-417(2001).
 CC -1- FUNCTION: Meiosis specific component of cohesin complex. The
 CC cohesin complex is required for the cohesion of sister chromatids
 CC after DNA replication. The cohesin complex apparently forms a
 CC large proteinaceous ring within which sister chromatids can be
 CC trapped. At anaphase, the complex is cleaved and dissociates from
 CC chromatin, allowing sister chromatids to segregate. The meiosis-
 CC specific cohesin complex probably replaces mitosis specific
 CC cohesin complex when it dissociates from chromatin during prophase
 CC I. (By similarity).
 CC -1- SUBUNIT: Component of the meiosis-specific cohesin complex, which
 CC also contains the SMC1 (SMC1L1 or SMC1L2) and SMC3 heterodimer.
 CC Such complex likely contains RAD21, or the meiosis-specific
 CC related protein REC8 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.
 CC In prophase I stage of meiosis, it is found along the axial
 CC elements of synaptonemal complexes. In late-pachytene-diplotene,
 CC the bulk of protein dissociates from the chromosome arms probably
 CC because of phosphorylation by PLK, except at centromeres, where
 CC cohesin complexes remain. It however remains chromatin associated
 CC at the centromeres up to metaphase I. During anaphase I, it
 CC probably dissociates from centromeres, allowing chromosomes

```

CC segregation (By similarity).
CC -!- TISSUE SPECIFICITY: Testis-specific.
CC -!- SIMILARITY: Belongs to the SCC3 family.
CC -!- SIMILARITY: Contains 1 SCD (stromalin conservative) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AY027880; AAK13052.1; -
DR InterPro: IPR005032; STAG.
DR Pfam: PF03365; STAG; 1.
KW Melosis; Cell cycle; Chromosome partition; Nuclear protein.
FT DOMAIN 324 409 SCD.
SQ SEQUENCE 1256 AA; 141950 MW; 1C0B3F49D48B85D CRC64;

Query Match          5.6%; Score 125.5; DB 1; Length 1256;
Best Local Similarity 37.2%; Pred. No. 3.3;
Matches 32; Conservative 13; Mismatches 24; Indels 17; Gaps 4;

Oy 2 QPQPGQGGQ---PQPGQGLGGGAGPAGGPGGPGPGPCLRELKLLLESIFPHRGHER 58
   ||: ||: ||| || || || || || || || || || || || || || || || ||
Db 1130 QSPRTVGRKRGKGGPGGP---GPGPGPGPGPGPGPGP-----ELICSQQLSGTQR 1180

Oy 59 FRIASACLDLSCEFLLAGAGAGAG 84
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 1181 LKMSAPGCFQIRCD-----PSGSLG 1201

RESULT 5
ID IF2_MYCTU
AC F71613;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR RV2839C OR MT2905 OR MTCY1687.03.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains",
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: One of the essential components for the initiation of

```

```

CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z81331; CAB03670.1; -
DR EMBL: AE007115; AAK47231.1; -
DR PIR: B70694; B70694.
DR HSP: P13551; IDAR.
DR TIGR: MT2905; -
DR Tuberculist; RV2839c; -
DR HAMAP; MF_00100; -; 1.
DR InterPro: IPR000795; EF_GTPbind.
DR InterPro: IPR004161; EFTU_D2.
DR InterPro: IPR000178; IF2.
DR InterPro: IPR006847; IF2_N.
DR InterPro: IPR001806; Ras_transfmrng.
DR InterPro: IPR005225; Small_GTP.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 2.
DR Pfam; PF04760; IF2_N; 2.
DR PRINTS; PR00315; ELONGATNFCT.
DR PRINTS; PR00449; RASTRNSFRMG.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMS; TIGR00487; IF-2; 1.
DR TIGRFAMS; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; FALSE_NEG.
DR KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 399 551 G-DOMAIN.
FT NP_BIND 405 412 GTP (BY SIMILARITY).
FT NP_BIND 455 459 GTP (BY SIMILARITY).
FT NP_BIND 509 512 GTP (BY SIMILARITY).
FT DOMAIN 250 255 POLY-GLY.
SQ SEQUENCE 900 AA; 94040 MW; F1B2F3C8A86952C2 CRC64;

Query Match          5.5%; Score 123.5; DB 1; Length 900;
Best Local Similarity 20.5%; Pred. No. 2.8;
Matches 92; Conservative 63; Mismatches 128; Indels 165; Gaps 21;

Oy 4 PQPGQGGQ-QPGPGQ-----QLGGGGAAPGAG--GGPGGPGPGPCLRELKLLLESIF 52
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 211 PRGGRPGAGAGRSDAGGNGYRGVGAPGTGFRGPGGGGGRPGQR-----260

Oy 53 HRGHERFRIASACLDLSCEFLLAGAGGAGAAAGPPLPPRG-----95
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 261 -----GGAGAGFRPG--GAPRRGKSKRKRQRYDSNQAP 294

Oy 96 -----SVP---GDPVRTHCNITESYPVPPVWSVSDPNLAALRLVDIKKNTLLQ 147
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 295 VVGVRLLPHNGETIRLARGASLSDF-----DKIDANPAALVQLNIGE---MVT 343

Oy 148 HLKRIISDLCKLYNLPHQPDVEMLDQPLPAEQCTQEDVSEDEEPEDELDHYEMK 207
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 344 ATQSGVDETLELGGSENNYVQV-----VSPEDRELE-SFDLSY---384

Oy 208 EEPFAEGKKSDEGGIGKENAI-----LEKIKK-NQKQDYNG-----244
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 385 -----GEDEG-GEEDLQVRPPVYVMGHVDHGKTRLLDTRKANVREAGAGTQH 434

Oy 245 -----AVSGSVQATDRLM-----KELRDIYRSQSFKGGNYAVELVNDLSYDNVLLKV 293
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db 435 IGAYQVAVDLGGSORLITFDTPGHEAFTAMRARGAKATDIAI-----LVVA 481

```



```

DR ProDom; PD000461; UBQ_conjugat; 1.
DR SMART; SM00212; UBCC; 1.
DR PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
DR PROSITE; PS00127; UBIQUITIN_CONJUGAT_2; 1.
KW ubl conjugation pathway; Ligase; Multigene family; Mitosis;
KW Cell cycle; Cell division; 3D-structure.
FT BINDING 114 114 UBIQUITIN.
FT MUTAGEN 114 114 C->S: INHIBITION OF CYCLIN B DEGRADATION.
FT HELIX 31 45
FT TURN 48 54
FT STRAND 50 54
FT TURN 59 60
FT STRAND 61 67
FT TURN 70 71
FT STRAND 73 76
FT TURN 78 84
FT STRAND 87 90
FT TURN 95 98
FT STRAND 104 104
FT TURN 105 106
FT STRAND 107 107
FT TURN 109 110
FT STRAND 112 113
FT TURN 116 118
FT HELIX 119 121
FT TURN 124 125
FT STRAND 128 138
FT TURN 139 140
FT HELIX 150 155
FT TURN 156 157
FT HELIX 159 175
SQ SEQUENCE 177 AA; 20134 MW; 06A454E798EE4F3A CRC64;

Query Match
Best Local Similarity 5.5%; Score 122.5; DB 1; Length 177;
Matches 45; Conservative 35; Mismatches 70; Indels 33; Gaps 10;

QY 234 KKNQDYLVGAVSGVQATRLMKELDIYRS-----QSPKGGNYAVELVNDLSYDNV 288
Db 15 QKERPRDVTTSKERHSV--SKRLOELRTLMLSGDPGITAPDG-----DNLFKV-V 63
QY 289 KLLKVDQPSALHNDLIQILKEGADFIILNFSKDNFPDPFVVRVSPVLSGGVVLGGG 348
Db 64 ATLDPKQTVTES-----LTKYK-----LTLEFSDYDKPPVVRFTPCWHPN-VDQSG 111
QY 349 AICMELLRKQGWSSAYSIESVIMQISATLVKRGKARVQFGANKSQYSLTRAQSQYKSLVQI 408
Db 112 NICLDIL-KENWTASYDVRTLLSLQSL--GEPNNAAPLNAQAADMSNQTETKVV--L 166
QY 409 HEK 411
Db 167 HEK 169

RESULT 8
AGOL_ARATH STANDARD; PRT; 1048 AA.
AC 004379;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Argonaute protein.
GN AGOL OR ATIG48410 OR FLIA17.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Leaf;
RX MEDLINE=98090460; PubMed=9427751;
RA Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;

*AGOL defines a novel locus of Arabidopsis controlling leaf development.*;
EMBO J. 17:170-180(1998).
[2]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA White O., Alonso J., Altafi H., Araujo R., Federspiel N.A., Kaul S.,
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Federspiel N.A., Kaul S.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Convey A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
"Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
Nature 408:816-820(2000).
-|- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL ORGANS, AND FORMATION OF AXILLARY MERISTEMS.
-|- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
-|- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
-|- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL STAGES.
-|- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
-|- SIMILARITY: Contains 1 PAZ domain.
-|- SIMILARITY: Contains 1 Piwi domain.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
EMBL; U91995; AAC18440.1; -
EMBL; AC007932; AAD49755.1; -
InterPro; IPR003100; PAZ.
InterPro; IPR003165; Piwi.
Pfam; PF02170; PAZ; 1.
Pfam; PF02171; Piwi; 1.
PROSITE; PS50821; PAZ; 1.
PROSITE; PS50822; PIWI; 1.
Developmental protein.
DOMAIN 391 501 PAZ.
DOMAIN 676 997 PIWI.
DOMAIN 13 104 GLY-RICH.
SQ SEQUENCE 1048 AA; 116190 MW; 3E5146343A09C541 CRC64;

Query Match
Best Local Similarity 5.5%; Score 122.5; DB 1; Length 1048;
Matches 87; Conservative 46; Mismatches 137; Indels 171; Gaps 21;

QY 3 QPQPQGGQGPQGP-----GQQLGG-----QGAAPGAGGGGPGGPGPCLRLRLKLESIFHRG 55
Db 66 QPPOQQQYGGPQYQVGRGGPPHQGGRGYGGGGGSSGP----- 109
QY 56 HERFRIASACLDLSCFLLAGAGAGAGAPGPHLPGRGVPDVRHCHNITESPAPV 115
Db 110 -----PQRQSV-----ELQAT 122
QY 116 PPIWVSDDPNLAALVRLYDIKGNVTLQLHLKRI-----ISDLCKLYNL 162
Db 123 SPTYQAVSSQPTLSEVSTQV---PEPTVLAQFQELSVQEQAGPSAQIPIPSKAKFK 179

```

QY 163 PQHPDVMLOPLPAEQTOEDVSEDEEMPEDETDLDHYE--MKEEPAEGKKSDD 220
 DB 180 PWRP-----GKQSKRCI--VRANHFALPD--KOLHYDVITTEVTSRG----- 223
 QY 221 GIGKENLAILEKIKKNQDYLNG---AVSGSVQATDRMLKELDIYRS--QSPKGGNYA 275
 DB 224 ----VNRAVMQOLVDNYRDSHLSRLPAYDGR-----KSLYTAGPLPNSKEFR 268
 QY 276 VELVNDSL-----YDNVVKL-----LKVDQDSALHNDLQILKEGAD 313
 DB 269 INLLDEEVGAGGQRREREFVKIKLVARADLHLGMFLEKQSDAPQALQVL-----D 322
 QY 314 FILLNFSKDNFPDPVPRVSVPLSGYVGLGGAICMELLTKQGSAY-SIESVIMQ 372
 DB 323 IVLRPLPSRIPVGRSP--YSPDQKKGSLGDL-----ESWRFQYSIRPTOMG 371
 QY 373 ISATL-----VKGKARVQF 386
 DB 372 LSLNIDMSSTAFIEANPVIQF 392

RESULT 9

UBC6_MOUSE STANDARD: PRT; 193 AA.
 AC P52482;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin-conjugating enzyme E2 E1 (EC 6.3.2.19) (Ubiquitin-protein
 DE ligase E1) (Ubiquitin carrier protein E1) (Ubch6).
 GN UBE2E1 OR UBCH6 OR UBCM3 OR UBCE5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalón D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Fahey J., Heiton E., Kettingan M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
 CC PROTEINS. MEDIATES THE SELECTIVE DEGRADATION OF SHORT-LIVED AND
 CC ABNORMAL PROTEINS.
 CC -1- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +
 CC diphosphate + protein N-ubiquityllysine.

CC -1- PATHWAY: Ubiquitin conjugation; second step.
 CC -1- SUBUNIT: INTERACTS WITH RNFI4.
 CC -1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-
 CC thiolester formation (By similarity).
 CC -1- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.
 CC STRONGEST, TO DROSOPHILA UBCD2.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X92665; CAAG3353.1; -
 CC EMBL; BC003781; AAH03781.1; -
 CC HSSP; P15731; 10CO.
 CC MGD; MGI:107411; Ube2el.
 CC InterPro; IPR000608; UBQ_conjugat.
 CC Pfam; PF00179; UBQ_con; 1.
 CC PRODOM; PD000461; UBQ_conjugat; 1.
 CC SMART; SM00212; UBCc; 1.
 CC PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
 CC PROSITE; PS00127; UBIQUITIN_CONJUGAT_2; 1.
 CC Ubl conjugation pathway; Ligase; Multigene family.
 CC DOMAIN 9 18 POLY-SER.
 CC BINDING 131 131 UBIQUITIN (BY SIMILARITY).
 CC SEQUENCE 193 AA; 21333 MW; B535B3095EF6C445 CRC64;
 SQ
 Query Match 5.58; Score 121.5; DB 1; Length 193;
 Best Local Similarity 22.98; Pred. No. 0.48; Indels 37; Gaps 8;
 Matches 49; Conservative 33; Mismatches 95;
 QY 187 SEDDEEMPEDETDLDHYEMKEEPAEGKKSDDGIGKENLAILEKIKKNQDYLNGAV 246
 DB 2 SDDDSRASTSSSSSSNQTEKEGSTPKKESKYSKSKLLTSK----- 50
 QY 247 SGSVQATDRMLKELDIYRSQSPKGGNYAVELVNDSLYDNVVKLLKVDQDSALHNDLQIL 306
 DB 51 -----RIQKELADITLDP-----PNCSPKGGDIYEWRSITL--GPPGSVY----- 91
 QY 307 KEKEGADFILLNFSKDNFPDPVPRVSVPLSGYVGLGGAICMELLTKQGSAYSI 366
 DB 92 ---EGGVF-FLDITFTPEYFPKPKVTFRTRIYHCN-INSQGVICLDIL-KDNWSPALTI 145
 QY 367 ESMVIMQISATLVK-GKARVQFGANKSOYSILTRAQ 399
 DB 146 SKVLLSICSLTDCNPADPLVGSITATQYMTNRAE 179
 RESULT 10
 ID EFER_HUMAN STANDARD; PRT; 756 AA.
 AC Q75154; Q9H155; Q9H1G0; Q9NU010;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Eferin.
 GN KIAA0665.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21486406; PubMed=11481332;
 RA Prekeris R., Davies J.M., Scheller R.H.;
 RT "Identification of a novel Rab11/25 binding domain present in eferin
 RT and rip proteins."
 RL J. Biol. Chem. 276:38966-38970(2001).
 RN [2]
 RP SEQUENCE FROM N.A.

```

RC TISSUE-Brain;
RX MEDLINE-98403980; PubMed-9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.*;
RL DNA Res. 5:169-176(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-21096910; PubMed-11157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgins D.R.;
RT *Sequence, structure and pathology of the fully annotated terminal 2
RT Mb of the short arm of human chromosome 16.*;
RL Hum. Mol. Genet. 10:339-352(2001).
RN [4]
RP SEQUENCE FROM N.A.
RX Wallis J., Lloyd C., Hall R.;
RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: Binds to RAB11 and RAB25.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF395731; AAL12940.1; -
DR EMBL; AB014565; BAA31640.1; -
DR EMBL; AE006463; AAK61232.1; -
DR EMBL; AL023881; CAB92745.1; -
DR EMBL; AL049542; CAC17519.1; -
DR EMBL; Z98882; CAC17523.1; -
DR PIR; T00367; T00367.
DR GO; GO:0017137; F:RAB Interactor activity; NAS.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2.
DR SMART; SM00054; EFh; 2.
KW Calcium-binding; Repeat.
FT CA_BIND 215 226 EF-HAND 1 (POTENTIAL).
FT CA_BIND 247 258 EF-HAND 2 (POTENTIAL).
SQ SEQUENCE 756 AA; 82439 MW; 264CEC399F28AFB9 CRC64;

Query Match 5.4%; Score 120; DB 1; Length 756;
Best Local Similarity 23.8%; Pred. NO. 3.7;
Matches 63; Conservative 28; Mismatches 104; Indels 70; Gaps 13;

QY 4 POPOGQDPGPGQQLGG-----QGAAPGAGGSP-----GGGPGPG 38
Db 50 PPGSLDPAPGAADGARGSAGPAPGEGPDPGSPAPPPRPGRLASDPAPG 109
QY 39 PCLRELKLLSIFRHRHERIFRISACLDLSCEFLAGAGAGAGA-----APGPHLP 92
Db 110 PRSEAPLELDFSWTEPEECGPASCPPE-SAPPRLOGSSSHRACEVDVFSFP-AP 167
QY 93 PRGSV-----PGDPVRIHCNTITESPAVP--PIWSVESDDPNLAALVRLYDIKNGTLL 145
Db 168 TAGELALQGPGSPQ-PSDLSQTHP-LPSEFVGSQE-DGPRLRAVFDALDGDGDFVRI 224
QY 146 LQHLKRIISDLCKLYNLQHPDVMLDQPL-----PAQCQTED 184
Db 225 EDFIQ-----FATVYGAQVQKDLTKYLDPSGLGVISFEDFYOGITAINRGDPGCCYGV 279
QY 185 VSSEDEDE-EMPEDEFDLDHYEMKE 208
Db 280 ASAQDEELPACDFEDDFVTEANE 304

```

```

RESULT 11
GCFC_HUMAN STANDARD; PRT; 917 AA.
ID Q9Y5B6; Q9NTQ0;
AC Q9Y5B6; Q9NTQ0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE GC-rich sequence DNA-binding factor homolog.
GN GCFC OR C21ORF66.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D).
RX MEDLINE-21564202; PubMed-11170702;
RA Raymond A., Friedli M., Neergaard Henriksen C., Chapot F.,
RA Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,
RA Antonarakis S.E.;
RT *From PREDs and open reading frames to cDNA isolation: revisiting the
RT human chromosome 21 transcription map.*;
RL Genomics 78:46-54(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE-Brain, Heart, Kidney, Lung, Muscle, Placenta, and Testis;
RA Chappot-Skovgaard F.M., Guipponi M., Lyle R., Antonarakis S.E.;
RT *Isolation and initial characterization of a putative human chromosome
RT 21 transcription factor.*;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 132-917 FROM N.A. (ISOFORM A).
RX MEDLINE-20237674; PubMed-10773462;
RA Slavov D., Hattori M., Sakaki Y., Rosenthal A., Shimizu N.,
RA Minoshima S., Kudoh J., Yaspo M., Ramser J., Reinhardt R., Reimer C.,
RA Clancy K., Rynditch A., Gardiner K.;
RT *Criteria for gene identification and features of genome organization:
RT analysis of 6.5 Mb of DNA sequence from human chromosome 21.*;
RL Gene 247:215-232(2000).
RN [4]
RP SEQUENCE OF 489-815 FROM N.A. (ISOFORM B).
RA Teramoto T., Thorgerirsson S.S.;
RT Cloning of candidate of GC-rich sequence DNA-binding factor.*;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Possible transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Name=A;
CC IsoId=Q9Y5B6-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q9Y5B6-2; Sequence=VSP_004267;
CC Name=C;
CC IsoId=Q9Y5B6-3; Sequence=VSP_004263, VSP_004264;
CC Name=D;
CC IsoId=Q9Y5B6-4; Sequence=VSP_004265, VSP_004266;
CC Note=Major isoform;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: TO GCFTCP9.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY033903; AAK68721.1; -
DR EMBL; AY033904; AAK68722.1; -
DR EMBL; AY033905; AAK68723.1; -
DR EMBL; AY033906; AAK68724.1; -
DR EMBL; AJ279080; CAC40813.1; -
DR EMBL; AF231920; AAF72944.1; -

```

```
DR EMBL; AF153208; AAD34617.1; ALT_TERM.
DR Genew; HGNC:13579; C21orf66.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003677; F:DNA binding activity; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR Transcription regulation; DNA-binding; Nuclear protein;
DR Alternative splicing.
DR VARSPLIC 462 469 VPLINELE -> SVQFRKLL (in isoform C).
FT VARSPLIC 470 917 /FTid-VSP_004263.
FT VARSPLIC 503 510 /FTid-VSP_004264.
FT VARSPLIC 511 917 /FTid-VSP_004265.
FT VARSPLIC 779 917 /FTid-VSP_004266.
FT VARSPLIC 799 917 /FTid-VSP_004267.
FT CONFLICT 692 711 MISSING (IN REF. 4).
FT SEQUENCE 917 AA; 104803 MW; 855960A1D50A7789 CRC64;

Query Match 5.4%; Score 120; DB 1; Length 917;
Best Local Similarity 23.3%; Pred. No. 4.7;
Matches 80; Conservative 38; Mismatches 115; Indels 110; Gaps 18;

Qy 9 QQQPGGQQLGGGAAPGAGG---PGG-----GPGPGCLRLKLLKLLSIFHRGHERFRI 61
Db 26 EQEPPPLPPGTGEEAGPGGDRAPGESLPGPSP----- 63
Qy 62 ASACLDLSCFLAGAGGAGAGPGLPGRSPGDPVRIHCNTITESVPVPTWSV 121
Db 64 PSALTPLGAB-----AGGFPGGAEPGNGLKPR-KRPRE-----NKEVPRASLLSFQ 110
Qy 122 ESDDPNLAALVRLVDIKGMNTLLQLHLKRIISDLCKLYNLPHQHPDEM----- 170
Db 111 DEEEN-----EEVFKVKSS-----YSKIVKLLKREYK-----EDLEKSKIKELNLSA 156
Qy 171 -LDQPLP-----AEQCTQED--VSSEDEDEEMPEDELDHYEMKEEPAEGKKSSEDGIG 223
Db 157 ESEQPLDKTHVKDTNQEDGVIISEHGDEDMSE-----KEEE-----KPKTGGAF 204
Qy 224 KENLAILEKIKKRODYLVANGSVSQVATDLAKELDIYRSQFKG-GNYAVELYNDS 282
Db 205 SNALSSNLVLRGPIPD---AAFTHAARKKQMARLGDFTPHDNEPGKGLVREDEDA 261
Qy 283 LYDNVKKLLKVDPSALHNDLQILKEK-----EGAD 313
Db 262 SDD-----EDDEKRRIVSVKESQKQIAEIGIEGSD 296

RESULT 12
ID UBC6_HUMAN STANDARD; PRT; 193 AA.
AC P51965;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ubiquitin-conjugating enzyme E2 E1 (EC 6.3.2.19) (Ubiquitin-protein
DE ligase E1) (Ubiquitin carrier protein E1) (Ubch6).
DE UBE2E1 OR UBC6.
CN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96162027; Pubmed=8576257;
RA Nuber U., Schwarz S., Kaiser P., Schneider R., Scheffner M.;
RT "Cloning of human ubiquitin-conjugating enzymes Ubch6 and Ubch7
RT (E2-F1) and characterization of their interaction with E6-AP and
```

```
RT RSP5 ";
RL J. Biol. Chem. 271:2795-2800(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Colon;
RX MEDLINE=22388257; Pubmed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -I- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS. MEDIATES THE SELECTIVE DEGRADATION OF SHORT-LIVED AND
CC ABNORMAL PROTEINS.
CC -I- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +
CC diphosphate + protein N-ubiquityllysine.
CC -I- PATHWAY: Ubiquitin conjugation; second step.
CC -I- SUBUNIT: INTERACTS WITH RNFI4.
CC -I- MISCELLANEOUS: A cysteine residue is required for ubiquitin-
CC thiolester formation.
CC -I- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.
CC -I- SOURCE: TO DROSOPHILA UBCD2
CC -I- SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -I- EMBL; X92963; CAA63539.1; -
CC EMBL; BC009139; AAH09139.1; -
CC HSP; P15731; I0CQ.
CC Genew; HGNC:12477; UBE2E1.
CC MIM; 602916; -
CC GO; GO:0004840; F:ubiquitin conjugating enzyme activity; TAS.
CC GO; GO:0004842; F:ubiquitin-protein ligase activity; TAS.
CC InterPro; IPR000608; UBU_conjugat.
CC Pfam; PF00179; UQ_con; 1.
CC ProDom; PD000461; UBU_conjugat; 1.
CC SMART; SM00212; UBC; 1.
CC PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
CC PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.
CC Ubl conjugation pathway; Ligase; Multigene family.
FT DOMAIN 9 18 POLY-SER.
FT BINDING 131 131 UBIQUITIN (BY SIMILARITY).
FT SEQUENCE 193 AA; 21404 MW; 2FBC50BE2A6A0008 CRC64;

Query Match 5.4%; Score 119.5; DB 1; Length 193;
Best Local Similarity 22.9%; Pred. No. 0.64;
Matches 49; Conservative 32; Mismatches 96; Indels 37; Gaps 8;

Qy 187 SEDEDEEMPEDELDHYEMKEEPAEGKKSDDGIGKENLAILEKIKKRODYLVANGV 246
Db 2 SDDDSRASTSSSSSSSSSQQTETPTPKKESKVSMSKSKLLTSK----- 50
Qy 247 SGVQVATDLAKELDIYRSQFKGNYAVELYNDSLYDNVKKLLKVDPSALHNDLQIL 306
```

Db 51 -----RIQELADITLDP--PNC-SAGPKGDNIEYEWSTIL--GPPG-SVY----- 91
 QY 307 KEKEGADFTLLNFSKDPFPDPVPRVYVSPVLGGYVLGGGAIACMELLTKQWSSAYSII 366
 Db 92 ---EGVF-FLDITTPFPKPKVTRTRIHCN-INSQVICLDIL-KDNWSPALTI 145
 QY 367 ESMVOISATLVK-OKARVQFGANKSOYSLTRAQ 399
 Db 146 SKVLLSICSLTDCNPADPLVGSITATQYMTNRAE 179

RESULT 13
 BAT8_HUMAN
 ID BAT8_HUMAN STANDARD; PRT: 1210 AA.
 AC Q96KQ7; Q14349; Q96MH5; Q96QD0; Q9UQL8; Q9Y331;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Histone-lysine N-methyltransferase, H3 lysine-9 specific 3
 DE (EC 2.1.1.43) (Histone H3-K9 methyltransferase 3) (H3-K9-HMTase 3)
 DE (HLA-B associated transcript 8) (G9a) (NG36).
 GN BAT8 OR G9A OR NG36.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Qin S., Madan A., Dickhoff R., Dors M., Madan A., Hicks P.,
 RA Loretz C., Ratcliffe A., Abbasi N., Shaffer T., Hood L.;
 RT "Sequence of the human major histocompatibility complex class III
 RT region".
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hirakawa M., Yamaguchi H., Imai K., Shimada J., Shina S., Tamiya G.,
 RA Oka A., Inoko H.;
 RT "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING (ISOFORM 2), AND
 RP TISSUE SPECIFICITY.
 RX MEDLINE=213564388; PubMed=11707778;
 RA Brown S.E., Campbell R.D., Sanderson C.M.;
 RT "Novel NG36/G9a gene products encoded within the human and mouse MHC
 RT class III regions".
 RL Mamm. Genome 12:916-924(2001).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Salivary gland;
 RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
 RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Katui-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Murakawa K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
 RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
 RT "NEO human cDNA sequencing project".
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 5-1210 FROM N.A. (ISOFORM 1).
 RC TISSUE=Muscle, and Uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield A.S., Krzywinski M.I., Jones S.J.M., Marra M.A.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [6]
 RP SEQUENCE OF 195-1210 FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Histocytic lymphoma;
 RX MEDLINE=93207535; PubMed=8457211;
 RA Milner C.M., Campbell R.D.;
 RT "The G9a gene in the human major histocompatibility complex encodes a
 RT novel protein containing ankyrin-like repeats".
 RL Biochem. J. 290:811-818(1993).
 RN [7]
 RP ENZYMIC ACTIVITY, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21326082; PubMed=11316813;
 RA Tachibana M., Sugimoto K., Fukushima T., Shinkai Y.;
 RT "Set domain-containing protein, G9a, is a novel lysine-preferring
 RT mammalian histone methyltransferase with hyperactivity and specific
 RT selectivity to lysines 9 and 27 of histone H3".
 RL J. Biol. Chem. 276:25309-25317(2001).
 RN [8]
 RP IDENTIFICATION IN COMPLEX WITH E2F6; TDFP1; MAX; MGA; EHM7ASE1; CBX3;
 RP RING1; RNF2; MBLR; L3MBTL2 AND YAP2.
 RX MEDLINE=21999559; PubMed=12004135;
 RA Ogawa H., Ishiguro K.-I., Gaubatz S., Livingston D.M., Nakatani Y.;
 RT "A complex with chromatin modifiers that occupies E2F- and
 RT Myc-responsive genes in G0 cells".
 RL Science 296:1132-1136(2002).
 CC -1- FUNCTION: Histone methyltransferase. Preferentially methylates
 CC Lys-9 of histone H3 and Lys-27 of histone H3 (in vitro). H3 Lys-9
 CC methylation represents a specific tag for epigenetic
 CC transcriptional repression by recruiting H3 proteins to
 CC methylated histones. Probably targeted to histone H3 by different
 CC DNA-binding proteins like E2F6, MGA, MAX and/or DPL. Also
 CC methylates histone H1 (By similarity).
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
 CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
 CC -1- SUBUNIT: Part of the E2F6-com-1 complex in G0 phase composed of
 CC E2F6, MGA, MAX, TDFP1, CBX3, BAT8, EHM7ASE1, RING1, RNF2, MBLR,
 CC L3MBTL2 and YAP2.
 CC -1- SUBCELLULAR LOCATION: Nuclear; associates with euchromatic
 CC regions. Does not associate with heterochromatin.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing: Named isoforms-3;
 CC Comment-Additional isoforms seem to exist;
 CC Name-1;
 CC IsoId=Q96KQ7-1; Sequence=Displayed;
 CC Name-2; Synonyms=NG36G9a-SPI;
 CC IsoId=Q96KQ7-2; Sequence=VSP_002211;
 CC Name-3; Synonyms=NG36;
 CC IsoId=Q96KQ7-3; Sequence=VSP_002212, VSP_002213;
 CC -1- TISSUE SPECIFICITY: Expressed in all tissues examined, with high
 CC levels in fetal liver, thymus, lymph node, spleen and peripheral
 CC blood leukocytes and lower level in bone marrow.
 CC -1- SIMILARITY: BELONGS TO THE HISTONE-LYSINE METHYLTRANSFERASE
 CC FAMILY. SUBVAR3-9 SUBFAMILY.
 CC -1- SIMILARITY: Contains 7 ANK repeats.
 CC -1- SIMILARITY: Contains 1 pre-SET domain.
 CC -1- SIMILARITY: Contains 1 SET domain.
 CC -1- SIMILARITY: Contains 1 post-SET domain.
 CC -1- CAUTION: NG36 and G9a were originally thought to derive from two
 CC separate genes. Ref.3 shows that all G9a transcripts also contain
 CC the in frame coding sequence of NG36.
 CC -1- CAUTION: Ref.1 and Ref.2 sequences differ from that shown due to
 CC erroneous gene model prediction.
 CC -1- CAUTION: It is not known whether Met-1 or Met-21 is the initiator


```

DR PIR; S12746; S12746.
DR HSP; Q54997; 1BL8.
DR FlyBase; FBgn003383; Shab.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR005821; Ion_trans.
DR InterPro; IPR001822; K_channel_pore.
DR InterPro; IPR003091; K_channel.
DR InterPro; IPR003131; K_tetra.
DR InterPro; IPR003971; Kv9_channel.
DR InterPro; IPR003968; Kv_channel.
DR InterPro; IPR005820; M_channel_nlg.
DR Pfam; PF00520; Ion_trans; 1.
DR Pfam; PF02214; K_tetra; 1.
DR PRINTS; PR00169; KCHANNEL.
DR PRINTS; PR01494; KV9CHANNEL.
DR PRINTS; PR01491; KVCHANNEL.
DR SMART; SM00225; BTB; 1.
KW Transport; Ion transport; Ionic channel; Voltage-gated channel;
KW Potassium channel; Potassium transport; Potassium; Transmembrane;
KW Glycoprotein; Phosphorylation; Multigene family; Alternative splicing.
FT TRANSMEM 436 454 SEGMENT S1.
FT TRANSMEM 474 495 SEGMENT S2.
FT TRANSMEM 506 527 SEGMENT S3.
FT TRANSMEM 536 561 SEGMENT S4.
FT TRANSMEM 577 598 SEGMENT S5.
FT TRANSMEM 638 659 SEGMENT S6.
FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 429 429 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 749 749 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 888 888 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).
FT MOD_RES 690 690 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT MOD_RES 731 731 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT MOD_RES 796 796 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT VARSPLIC 717 746 Missing (in isoform Short).
FT FTid-VSP_000960.
FT R->: IN ALLELE SHAB-1; TEMPERATURE-SENSITIVE PARALYTIC.
FT V->: IN ALLELE SHAB-1; TEMPERATURE-SENSITIVE PARALYTIC.
FT Q -> L (IN REF. 1, 2 AND 3).
FT S -> G (IN REF. 1, 2 AND 3).
FT R -> G (IN REF. 1, 2 AND 3).
FT S -> G (IN REF. 1, 2 AND 3).
FT FS -> CA (IN REF. 5).
FT A -> S (IN REF. 1, 2 AND 3).
FT V -> I (IN REF. 5).
FT A -> T (IN REF. 5).
FT C -> Y (IN REF. 5).
FT V -> I (IN REF. 5).
FT EQ -> DE (IN REF. 1, 2 AND 3).
FT A -> G (IN REF. 1, 2 AND 3).
FT GDGGGVDNDINLQAKGLPIQMHTPGVEAEIIRROVALENLQNRMDLQDQVPEFCFCFTKGLPGHCICPLRANSV -> VMEMGAVSMTTTPRPPDCPSR (IN REF. 1, 2 AND 3).
FT DDD -> NDN (IN REF. 4).
FT QY 13 GFGQGLGGGAAPGAGGGPGPGP-CLRRRLKLESIFRHGRHFRFRIASCLD-ELS 70
Db 163 GAGASVTCGSGAGTGTGTGAGSGSGSAGKQVR-----YAPFPVAPSTHSIPT 213
QY 71 CEFLLAGAGAGAGAAPGPHLPGRGVP-----GDPVRIHCNITESYPVPPITWSE 122
Query Match 5.3%; Score 118; DB 1; Length 985;
Best Local Similarity 23.5%; Pred. No. 6.9;
Matches 63; Conservative 27; Mismatches 102; Indels 76; Gaps 12
QY 13 GFGQGLGGGAAPGAGGGPGPGP-CLRRRLKLESIFRHGRHFRFRIASCLD-ELS 70
Db 163 GAGASVTCGSGAGTGTGTGAGSGSGSAGKQVR-----YAPFPVAPSTHSIPT 213
QY 71 CEFLLAGAGAGAGAAPGPHLPGRGVP-----GDPVRIHCNITESYPVPPITWSE 122
Db 214 SQIIVSGYGVGVGGAGSSQSI--SGGVYTHSQSNITGALOTHSRMSKSIPTTPEPMIAQ 271

```


Search completed: July 29, 2003, 13:45:22
Job time : 25 secs

[illegible]

Db 28 SFHIELEDDSNFTWNTGVMVLNEDSIYH-----GGFFKAQMRFPDPFSPQ 76
Qy 332 VRVSPVLGGYVLGGGAICMELTKOG-----WSSAYSIESVIMQISATL--- 377
Db 77 FR-FTPAIYHPNVYRDGRCLISILHOSGDPMTDEPDAETWSPVQTVESVLISVSLLEDP 135
Qy 378 -----VKGKARVQFANKSQYS-----LTRAQOQSYKSLVQIHE--- 410
Db 136 NINSPANVDAADYRKNPQYKQVKMEVERSKODIPKGFIMPTSESAYISQSKLDEPES 195
Qy 411 -----KNGWYTPPKED 421
Db 196 NKDMADNEWYDSLDLD 211

RESULT 2

US-09-196-525-5

; Sequence 5, Application US/09196525
; Patent No. 5989883

GENERAL INFORMATION:

; APPLICANT: Au-Young, Janice

; APPLICANT: Goli, Surya K.

; APPLICANT: Hillman, Jennifer L.

; TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING

; TITLE OF INVENTION: ENZYME

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: U.S.

; ZIP: 94304

COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/196,525

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/679,765

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0093 US

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 295 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 788905

US-09-196-525-5

Query Match

Best Local Similarity 6.1%; Score 135.5; DB 2; Length 295;

Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;

Qy 273 NYAVELVNDLS-LYDNNVLLKVDODSALHNDLQILKEGADFILLNFSKDNFPDPF 331

Db 28 SFHIELEDDSNFTWNTGVMVLNEDSIYH-----GGFFKAQMRFPDPFSPQ 76

Qy 332 VRVSPVLGGYVLGGGAICMELTKOG-----WSSAYSIESVIMQISATL--- 377

Db 28 SFHIELEDDSNFTWNTGVMVLNEDSIYH-----GGFFKAQMRFPDPFSPQ 76

Db 77 FR-FTPAIYHPNVYRDGRCLISILHOSGDPMTDEPDAETWSPVQTVESVLISVSLLEDP 135
Qy 378 -----VKGKARVQFANKSQYS-----LTRAQOQSYKSLVQIHE--- 410
Db 136 NINSPANVDAADYRKNPQYKQVKMEVERSKODIPKGFIMPTSESAYISQSKLDEPES 195
Qy 411 -----KNGWYTPPKED 421
Db 196 NKDMADNEWYDSLDLD 211

RESULT 3

US-09-318-317-5

; Sequence 5, Application US/09318317

; Patent No. 6172199

GENERAL INFORMATION:

; APPLICANT: Au-Young, Janice

; APPLICANT: Goli, Surya K.

; APPLICANT: Hillman, Jennifer L.

; TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING

; TITLE OF INVENTION: ENZYME

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: U.S.

; ZIP: 94304

COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/318,317

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/196,525

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0093 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 295 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 788905

US-09-318-317-5

Query Match

Best Local Similarity 6.1%; Score 135.5; DB 3; Length 295;

Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;

Qy 273 NYAVELVNDLS-LYDNNVLLKVDODSALHNDLQILKEGADFILLNFSKDNFPDPF 331

Db 28 SFHIELEDDSNFTWNTGVMVLNEDSIYH-----GGFFKAQMRFPDPFSPQ 76

Qy 332 VRVSPVLGGYVLGGGAICMELTKOG-----WSSAYSIESVIMQISATL--- 377

Db 77 FR-FTPAIYHPNVYRDGRCLISILHOSGDPMTDEPDAETWSPVQTVESVLISVSLLEDP 135

Qy 378 -----VKGKARVQFANKSQYS-----LTRAQOQSYKSLVQIHE--- 410

Db 136 NINSPANVDAADVKNPEQYKQVKMEVERSKODIPKGFIMPTSESAYISQSKLDEPES 195
QY 411 -----KNGWYTPPKED 421
Db 196 NKDMADNFWYDSLDLD 211

RESULT 4

US-09-177-165A-22
; Sequence 22, Application US/09177165A
; Patent No. 6426205
; GENERAL INFORMATION:
; APPLICANT: Tvers, Mike
; APPLICANT: Willems, Andrew
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
; TITLE OF INVENTION: DEPENDENT PROTEOLYSIS
; FILE REFERENCE: 11757.10USU1
; CURRENT APPLICATION NUMBER: US/09/177.165A
; CURRENT FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/092,443
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/063,254
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 22
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-22

Query Match 6.1%; Score 135.5; DB 4; Length 295;
Best Local Similarity 21.4%; Pred. No. 0.0012;
Matches 42; Conservative 33; Mismatches 62; Indels 59; Gaps 7;
QY 273 NYAVELVNDLS-LYDWNVKKLVQDQDSALHNDLQILKEGADFILLNFSKDNFPDPFP 331
Db 28 SFHIELEDDSNIFWNGVWLNEDSIYH-----CGFFKAQMRFPEDFPFPQ 76
QY 332 VRVSPVLGGYVGGGGAICMELITKOG-----WSSAYSIESVIMQISATL--- 377
Db 77 FR-FTPAIYHPNVYRDGRICISILHQSGDPMTPDEDAETWSPVQTVESVLISVSLLED 135
QY 378 -----VKGKARVQFCANKSQYS-----LTRAQOSYKSLVQIHE--- 410
Db 136 NINSPANVDAADVKNPEQYKQVKMEVERSKODIPKGFIMPTSESAYISQSKLDEPES 195
QY 411 -----KNGWYTPPKED 421
Db 196 NKDMADNFWYDSLDLD 211

RESULT 5

US-08-828-533-3
; Sequence 3, Application US/08828533
; Patent No. 6180379
; GENERAL INFORMATION:
; APPLICANT: Ruderman, Joan V.
; APPLICANT: Hershko, Avram
; APPLICANT: Kirschner, Marc W.
; APPLICANT: Townsley, Fiona
; APPLICANT: Aristarkov, Alexander
; APPLICANT: Eytan, Esther
; APPLICANT: Yu, Hongtao
; TITLE OF INVENTION: NOVEL CYCLIN-SELECTIVE UBIQUITIN CARRIER
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,533
FILING DATE: 31-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kerner Ph.D., Ann-Louise
REGISTRATION NUMBER: 33,523
REFERENCE/DOCKET NUMBER: HAZ-01SCIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)526-6000
TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 178 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: linear
FRAGMENT TYPE: linear
US-08-828-533-3

Query Match 5.5%; Score 122.5; DB 3; Length 178;
Best Local Similarity 24.6%; Pred. No. 0.0076;
Matches 45; Conservative 35; Mismatches 70; Indels 33; Gaps 10;
QY 234 KKNQDYLNGAVSGSVQATDRLMKELRDIYS-----QSFKGGNVAVELVNDLSLDWNY 288
Db 15 QKERPRDMTTSKERHSV--SKRLQQLRLTLLMSGDCITAFPDG-----DNLEKVV 63
QY 289 KLLKVQDSDALHNDLQILKEGADFILLNFSKDNFPDPFPVRVSPVLGGYVGGG 348
Db 64 ATLDGPKDVTYES-----LKYK-----LTFLEFSDYIPKPPVVKVTTTCWHFN-VDQSG 111
QY 349 AICMELITKOGWSSAYSIESVIMQISATLVKGRVQFGANKSQYSILTRAQOSYKSLVQI 408
Db 112 NICLDIL-KENWTASYDVRTILLSLQSL--GEPNNASPLNAQAADWMSQTEYKVV--L 166
QY 409 HEK 411
Db 167 HEK 169

RESULT 6

US-09-772-156-3
; Sequence 3, Application US/09772156
; Patent No. 6528633
; GENERAL INFORMATION:
; APPLICANT: Ruderman, Joan V.
; APPLICANT: Hershko, Avram
; APPLICANT: Kirschner, Marc W.
; APPLICANT: Townsley, Fiona
; APPLICANT: Aristarkov, Alexander
; APPLICANT: Eytan, Esther
; APPLICANT: Yu, Hongtao
; TITLE OF INVENTION: NOVEL CYCLIN-SELECTIVE UBIQUITIN CARRIER
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/772,156
; FILING DATE: 31-Mar-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,533
; FILING DATE: 1997-03-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner Ph.D., Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: HAZ-015CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 178 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: <Unknown>
; FRAGMENT TYPE: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
;
US-09-772-156-3

```

```

Query Match 5.5%; Score 122.5; DB 4; Length 178;
Best Local Similarity 24.6%; Pred. No. 0.0076;
Matches 45; Conservative 35; Mismatches 70; Indels 33; Gaps 10;

QY 234 KKNQDYLVNGAVSGVQATRLMKELRDIVRS-----QSFKGGNYAVELVNDLSYDNVY 288
Db 15 QKERPRDWTTSKERHSV--SKRLQELRTLLMSGDPGITAPFDG-----DNLFKW-V 63

QY 289 KILKVDQDSALHNDLQILKEGADFILLNFSKDNFPDPFVVRVSPVLSGGVVLGG 348
Db 64 ATLDPGKDTVYES-----LKIK-----LTLEFPSDYPIKPPVVKFTPCWHPN-VDOSG 111

QY 349 AICMELLKQGWSSAYSIESVIMOISATLVKGRVQFGANKSQSLTRAQOQSKSLVQI 408
Db 112 NICLDIL-KENTASYDVVRTILLSLQSL--GEPNAPLNQAADMNSQTEYKKV--L 166

QY 409 HEK 411
Db 167 HEK 169

```

```

RESULT 7
US-07-720-589-2
; Sequence 2, Application US/07720589
; Patent No. 5324630
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Rance B.
; APPLICANT: Peking, Guey-Chen
; TITLE OF INVENTION: Methods and Compositions for Diagnosing
; TITLE OF INVENTION: Lyme Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James M. Heslin
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/720,589
; FILING DATE: 19910628

```

```

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-720-589-2

Query Match 5.4%; Score 120.5; DB 1; Length 700;
Best Local Similarity 21.3%; Pred. No. 0.085;
Matches 80; Conservative 62; Mismatches 133; Indels 101; Gaps 18;

QY 90 HLPFRGVPDPPVRIHCNITESYPAPVPIWSVESDDPNLAALVRLVYDIKKGNTLLQLHL 149
Db 339 NLPKPGDVSSPKVDKQLIKESLEDLQELK-ETGDENQKREIEKQIEIKKSEKLLK-- 395
QY 150 KRIISDLCKLYNLQHPDVEMLDQPLPAEQCTQEDVSEDE-----DEEMPEDETDLHY 204
Db 396 -----SKDDKASKDGKALDLDRELNSKASSKEKS 424
QY 205 EMKEEPAEGK--KSEDDGIGKENLAI-----LEKIKK-NQODYLVNGAVSGVQATDRL 256
Db 425 KAKEEITKGSQKSLGDLNNDENLMPEDQKLPVKKLDSKREF---KPVSEVEKLDKI 481
QY 257 MK-----ELRDIYRSQKGGNYAVELVNDLSYDNVVKLLKVDQDSALHNDLQILKEKE 310
Db 482 FKSNNVNGELSPLDKS-SYKIDDSKEETVKNQVNLQTKPQVKDQVTSLNEDLTMSIDS 540
QY 311 GADILLNFSKDNFPDPFVVRVSPVLSGGYV---LGGGATCMELLTKQGS--SAY 364
Db 541 SS-----PVFLEVIDPITNLGTQLDLNTG-VSLKESTQQGIQRYGIY 583
QY 365 SIESVIMOISATLVKGRVQF-----GANKSQSLTRAQOQSKS---LVOIHEK- 411
Db 584 EREKDLVVIK--WDSGRAKQLQILDLENKLVVSESNEFNKNSSLYVDSKMLYAVRDKD 641
QY 412 --NGW---YTPPKED 421
Db 642 SSNDWRLAKFSFKNLD 657

```

```

RESULT 8
US-08-785-190-2
; Sequence 2, Application US/08785190
; Patent No. 5977339
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Rance B.
; APPLICANT: Peking, Guey-Chen
; TITLE OF INVENTION: Methods and Compositions for Diagnosing
; TITLE OF INVENTION: Lyme Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James M. Heslin
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,190

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/05539
FILING DATE: 19920629
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-05539-2

```

Query Match	5.4%; Score 120.5; DB 5; Length 700;
Best Local Similarity	21.3%; Pred. NO. 0.085;
Matches	80; Conservative 62; Mismatches 133; Indels 101; Gaps 18;
QY	90 HLPGRGVCPVRIHCNITESYPAYPPVWSVSDDPNLAVALERLVDIKGNWLLLOHL 149
Db	339 NLPKPGDVSPKVDKQLKESLEDLQEQLK-ETGDENQKRREIKQEIKNKSEKLLK-- 395
QY	150 KRIISDLCKLYNLPQHPOVEMLDQPLPAEOCTQEDVSSE-----DEEMPEDEDLDHY 204
Db	396 -----SKDDKASKDCKALDDRELNSKASSKEKS 424
QY	205 EMKEEPPAGCK--KSEDDGIGKENLA-----LEKIK-NQROYLNGAVSGSVQATDRL 256
Db	425 KAKEEITTKGKSOKSLGDLNNDENLAMPEDQKLPEVKLDSKKEF--KPVSEVKLDKI 481
QY	257 MK-----ELRDLYRSQSFKGNYAVELYNDLSLDNNVLLKYVDQSALHNDLQILKEKE 310
Db	482 FKSNNVGBLSPLDKS-SYKDDSKSETVKNQKTPQVKDQVTSLSMEDLTTMSIDS 540
QY	311 GADFILLNFSFKDNFFDPFVRVSVFVLSGGYV-----LGGAICAICMELLTKQGS--SAY 364
Db	541 SS-----PVFLEVIDPITNLGTLQIDLNTG-VSLKESYQQGIQRYGIY 583
QY	365 STESVIMQISATILVAKCARVQF-----GANKSOYSLTRAQOOSYKS----LVQTHEK- 411
Db	584 EREKDLVVIK--MDSGAKQLILDKLENLKVVSSENFENKNSSLYDVKSMILVAVRDKD 641
QY	412 --NGW---YTPPKED 421
Db	642 SSNDWRLAKFSPKNLD 657

```

1 RESULT 10
2 US-08-679-765-3
3 US-08-679-765-3
4 : Sequence 3, Application US/08679765
5 : Patent No. 5840866
6 : GENERAL INFORMATION:
7 : APPLICANT: Au-Young, Janice
8 : APPLICANT: Goli, Surya K.
9 : APPLICANT: Hillman, Jennifer L.
10 : TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING ENZYME
11 : NUMBER OF SEQUENCES: 5
12 : CORRESPONDENCE ADDRESS:
13 : ADDRESSEE: Incyte Pharmaceuticals, Inc.
14 : STREET: 3174 Porter Drive
15 : CITY: Palo Alto
16 : STATE: CA
17 : COUNTRY: U.S.
18 : ZIP: 94304
19 : COMPUTER READABLE FORM:
20 : MEDIUM TYPE: Diskette
21 : COMPUTER: IBM Compatible
22

```



```

, , REFERENCE/DOCKET NUMBER: PF-0093 US
, ,
, , TELECOMMUNICATION INFORMATION:
, , TELEPHONE: 415-855-0555
, , TELEFAX: 415-845-4166
, , INFORMATION FOR SEQ ID NO: 3:
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 193 amino acids
, , TYPE: amino acid
, , STRANDEDNESS: single
, , TOPOLOGY: linear
, , MOLECULE TYPE: peptide
, , IMMEDIATE SOURCE:
, , LIBRARY: GenBank
, , CLONE: 1064914
, , US-09-318-317-3

```

Query Match	5.4%	Score 119.5	DB 3	Length 193
Best Local Similarity	22.9%	Pred. No. 0.015		
Matches	49	Conservative 32	Mismatches 96	Indels 37
Gaps	8			

QY	187	SEDEDEMPEDTEDLDHVMKEEPEAPGKKSSEDGICKENLAILEKIKKQRDYLN	GA	246
		: :		
Db	2	SDDDSRASTSSSSSSNQOTKETNTPPKKESKVSMSKNKLLSTSAK		50
QY	247	SGSVQATDRMLKELDIYRSQSKGGNYAYELVNDLSLYDNWKLKVDQDSALHNDLQIL		306
		: :		
Db	51	-----RIQKELADITLPP-----PNC\$AGPKGDNIEYK\$TIL-----GPF\$GVY-----		91
QY	307	KEKEGADFILLNF\$EKDNPFDPF\$YRVV\$FVL\$GGVYLVGGGACIMELLTKQGWSSNY\$		366
		: :		
Db	92	---EGGVF-FLDITTFPEYFK\$PKVTFTR\$YHCHN-INSQGVICLDLIL-KDNW\$PALTI-		145
QY	367	ESVIMQISAPLVK-GRARVOFGANK\$QY\$SLTRAQ		399
		: :		
Db	146	SKVLL\$IC\$SLTDCN\$PADPVL\$GV\$ATQYMTNRAE		179
		: :		

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: PGANNOT01
; CLONE: 61887
US-08-679-765-1

```

Query Match	5.3%	Score 119;	DB 2;	Length 201;
Best Local Similarity	21.1%	Pred. No. 0.018;		

QY	189	DEDEMPEDTDLDIHYEMKEEPEAGKKSDDGTGKENLAILEKIKKNQRQDYLVNGAVSG	248
Db	9	DSPSTSGSGSDGQRESVQGEPEP-----EQVOPKKKEGKISSKTA	51
QY	249	SVQ-ATDRLMKELRDIYSQSQFKGNAVYLVNDSLVDNVNKLKLVQDQDSALHNDLQILK	307
Db	52	KUSTAKRIQKELAEITLDP-----PNC SAGKGNIIYEWRSITL--GPPGSVY-----	99
QY	308	EKEGADFILLNFSFKDNEPDPFPVRYVSVLVGGYVLGGCAICMELLTKQGWSSAYSIE	367
Db	100	--EGGVF-FLDITFSPDPYFPKPKVTFTRIYHCN--INSQVVICLOIL-KDNWSPALITIS	154
QY	368	SVIMOISATLV-KGKARVQFGANKSQYSLTRAQ	399
Db	155	KVLISGICSLTDXNPADPLVGSIAQYMTNRAE	187

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

LIBRARY: PGANNOT01
CLONE: 61887
US-09-196-525-1

Query Match 5.3%; Score 119; DB 2; Length 201;

Best Local Similarity 21.1%; Pred. No. 0.018;
Matches 45; Conservative 42; Mismatches 90; Indels 36; Gaps 9;

QY 189 DEDEMPEDTDLHYEMKEEPAEGKSEDDGIGKENLALEKIKKNORODYINGAVSG 248
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 9 DDSFSTSGSGSDGQRESVQOEPR-----EQVQPKKKGKISSKTAA 51
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 249 SVQ-ATDRLMKELRDYRSQFKGKNYAVELVNDSLYDWNVKKLLKVDQDSALHNDLQILK 307
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 52 KLSTSAKRIQELAEITLDP--PNCAGPKGDNIEYWRSTIL--GPPGSVY----- 99
QY 308 EKEGADFILLNFSPKDNFPDPFVRVYSPVLGGYVLGGGAICMELLTKOGWSSAYSIE 367
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 100 --EGGVF-FLDITFSPDYFPKPKVTRTRIYHCN-INSQGVICLDIL-KDNWSPALTIS 154
QY 368 SVIMOISATLV-KGKARVOFGANKSOYSLTRAQ 399
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 155 KVLISCSLLTDXNPADPLVGSITATQYMTNRAE 187

RESULT 15

US-09-318-317-1
; Sequence 1, Application US/09318317
; Patent No. 6172199

; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING
; TITLE OF INVENTION: ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/318,317
; FILING DATE:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,525
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0093 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: PGANNOT01
; CLONE: 61887

US-09-318-317-1

Query Match 5.3%; Score 119; DB 3; Length 201;
Best Local Similarity 21.1%; Pred. No. 0.018;
Matches 45; Conservative 42; Mismatches 90; Indels 36; Gaps 9;
QY 189 DEDEMPEDTDLHYEMKEEPAEGKSEDDGIGKENLALEKIKKNORODYINGAVSG 248
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 9 DDSFSTSGSGSDGQRESVQOEPR-----EQVQPKKKGKISSKTAA 51
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 249 SVQ-ATDRLMKELRDYRSQFKGKNYAVELVNDSLYDWNVKKLLKVDQDSALHNDLQILK 307
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 52 KLSTSAKRIQELAEITLDP--PNCAGPKGDNIEYWRSTIL--GPPGSVY----- 99
QY 308 EKEGADFILLNFSPKDNFPDPFVRVYSPVLGGYVLGGGAICMELLTKOGWSSAYSIE 367
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 100 --EGGVF-FLDITFSPDYFPKPKVTRTRIYHCN-INSQGVICLDIL-KDNWSPALTIS 154
QY 368 SVIMOISATLV-KGKARVOFGANKSOYSLTRAQ 399
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 155 KVLISCSLLTDXNPADPLVGSITATQYMTNRAE 187

Search completed: July 29, 2003, 13:54:57
Job time : 32 secs